UNIVERSITY OF COPENHAGEN

Adoption of banana cultivation and information networks

An empirical study of northern rural Tanzania

Larsen, Anna Folke

Publication date:
2012

Document version
Early version, also known as pre-print

Citation for published version (APA):
Larsen, A. F. (2012). *Adoption of banana cultivation and information networks: An empirical study of northern rural Tanzania*. Kbh: Økonomisk institut, Københavns Universitet.

DET SAMFUNDSVIDENSKABELIGE FAKULTET
Økonomisk Institut
KØBENHAVNS UNIVERSITET

Kandidatspeciale

Anna Folke Larsen

Adoption of banana cultivation and information networks

An empirical study of northern rural Tanzania

Vejleder: Mette Ejrnæs

Antal ECTS: 30

Afleveret den: 27/06/12

Preface

This master thesis is a product of my collaboration with the Rockwool Foundation Research Unit (henceforth RFF) as a part of my 4+4 ph.d. programme. In the fall 2010 I worked as a consultant for RFF where my task was to design a quantitative impact evaluation of the Rockwool Initiative for Poverty Alleviation in Tanzania (RIPAT), an agricultural project in nortern rural Tanzania. Under the supervision of Helene Bie Lilleør, Head of Evaluation, I developed survey instruments, designed the survey and planned the execution of the data collection and data entry. In that connection, I visited Tanzania twice. First, a week in September to perform preliminary pilot-testing of questions for the questionnaires and to become acquainted with the area and the project. Second, five weeks in November-December together with Maria Fibæk, a master student at the Department of Economics, University of Copenhagen, where we prepared the data collection. The first week of this stay, we were accompanied by Helene Bie Lilleør.

The impact evaluation focussed on the impact of project participation on food security and poverty. As RIPAT was designed to stimulate diffusion of the new agricultural technologies, we chose to include a module in the survey pertaining diffusion and these data underlies the study at hand.

I would like to thank RFF for the financial support making this study possible. I am grateful to Helene for all guidance with respect to this study and for the close cooperation we have had. I would like to thank Maria for the indispensable support and cooperation during field work and Cathrine Søgaard Hansen for research assistance and supervision of data entry. Jens Vesterager, Programme Manager at the Rockwool Foundation, has provided me with very usefull inputs on RIPAT and different aspects of tropical agriculture which I appreciate. I would like to thank all staff at RECODA, the implementing organisation, for the cooperation and help, and in particular Catherine Maguzu, Deputy Head, for the many, many hours spend with me in the office and in the field and for being available for all kinds of questions. Finally, I thank my supervisor Mette Ejrnæs for invaluable guidance and comments.

All errors remain my own.

Contents

1 Introduction

Despite the fact that obesity is considered to be a global epidemic by the World Health Organisation, three out of ten people in low income countries still suffers from undernourishment.[1] Technological changes within the agricultural sector have the scope of ensuring the food security of subsistence farmers in developing countries, but the process of technological change is complex. The diffusion of new technologies is often slower than classical economic theory would predict, because farmers do not have perfect information about the new agricultural techniques and the relative profitability of a new technology may depend on local circumstances. The decision to adopt a new technology is affected by the farmer's uncertainty about its relative advantage, and when information flows from government and research institutions are limited or not trust-worthy, farmers will rely on information from neighbours and friends about new technologies.

Through information networks of the farmers, development projects with the aim of closing the technological gap may be able to increase their reach beyond the initial project participants. I study this hypothesis empirically using household data from northern rural Tanzania. A new banana cultivation technique was introduced by an agricultural project, RIPAT,[2] in 2006, which aimed at increasing the food security in the project area. I study farmers who did not initially sign up for the project and investigate to what extent discussing farming issues with participating farmers affects the decision to adopt improved banana cultivation. My research question is: How does adoption of improved banana cultivation among non-project participants depend on information links to the initial participating farmers?

To my knowledge, there exist no prior studies of such second generation project take-up in the literature of adoption of agricultural technologies. From a policy perspective, it is relevant to study the second generation take-up of a programme because continued diffusion of the introduced technology will increase the impact of the project. Under-

[1] The World Health Organisation states that obesity is a global epidemic here: http://www.who.int/nutrition/topics/obesity/en/. In 2008, 29 percent of the population in low income countries were undernourished according to the World Development Indicators: http://data.worldbank.org/data-catalog/world-development-indicators.

[2] RIPAT stands for the Rockwool Initiative for Poverty Alleviation in Tanzania.

standing the social learning processes in the dissemination of knowledge will have policy implications for efficient targeting of future project participants.

Within development economics, social learning has received increased attention over the past two decades. The theoretical literature on the subject grew rapidly in the 1990s (Banerjee, 1992; Besley and Case, 1993; Ellison and Fudenberg, 1993; Foster and Rosenzweig, 1995; Bardhan and Udry, 1999), whereas the empirical literature has followed more slowly in the later years (Munshi, 2004; Bandiera and Rasul, 2006; Conley and Udry, 2010). Munshi (2007) assigns these differential trends to the difficulty of identifying social learning effects. This paper will add to the thin empirical literature and identification will play a central role throughout the paper, guided by the pioneer study of identification of social interactions by Manski (1993).

The empirical analysis in this paper shows that farmers who discuss farming issues with banana growing project participants are more likely to adopt improved banana cultivation. Meanwhile there is also tentative evidence of an impact of discussing farming issues with banana growers who are not project participants, though this effect is substantially lower and less robust. These findings confirm the important role of information networks for the adoption decision as suggested by the main part of the theoretical literature and as found in the few other sound empirical studies. But in addition, this paper contributes to the literature by providing a policy suggestion for future agricultural projects focused on knowledge transfer: the results suggest that targeting well-connected farmers will increase the impact of the project through diffusion of technology in the local communities.

The data underlying this analysis were collecting in January 2011 for the purpose of an impact evaluation of the RIPAT project and for this study of diffusion of technology. The analysis is based on farmers who live in villages where the RIPAT project took place but who did not sign up for the project. Hence, I employ a subset of these data comprising of 520 households in eight rural villages in Arumeru district, Tanzania. My measure of the information network of the farmer has three components: The network size, the number of network members who grow improved bananas and among these, the RIPAT farmers in the network for whom I have detailed background information. All three network variables are captured with recall questions referring to the time before the decision to adopt was taken. Logit estimates show that for a given network size, the number of banana growers and in particular RIPAT banana growers in the network has a strong and significant positive effect on the propensity to adopt.[3]

The identification strategy relies on three falsification tests related to reverse causality and two different types of confounding effects that cause a positive correlation between adoption and the number of adopters in network. As a start, I consider the concern of *reverse causality*: despite the fact that networks are captured by recall questions, a farmer's ex ante perception of the profitability of banana cultivation may affect network formation before adoption. I exploit the presence of super farmers in the project (farmers who are trained to teach banana cultivation to peers), who would naturally be the ones to contact if a farmer is interested in banana cultivation. If they are better represented in the network of non-participating farmers than regular RIPAT farmers are, that could indicate that the network measure is endogenous. I show that super farmers are not more likely to occur in the network of a non-participating farmer than a regular RIPAT farmer is, which indicates that my network measure is not endogenous to the adoption decision. To proceed, I employ the useful distinction of Manski (1993) between endogenous social interaction effects, contextual effects and correlated effects. First, I address *contextual effects* which are defined as the impact of the characteristics of the network on the individual adoption decision. They are also considered as social interaction effects, but (Manski, 1993, p.:532) term them "exogenous" because they relate to the effect of the *characteristics* of network members, such as education and wealth, rather than the impact of network adoption *behaviour*. As I only have elaborate data on the RIPAT farmers in the network I address contextual effects by splitting the RIPAT farmers in the network according to socioeconomic characteristics to investigate if the network effect is heterogeneous across characteristics of network members. I do not find support for heterogeneous effects which indicates that network effects are not driven by contextual effects. Second, *correlated effects* represent factors that induce a correlation in behaviour within networks

[3] I consider *improved* banana cultivation, but for simplicity I disregard the word "improved". I will always state explicitly if I refer to *traditional* banana cultivation.

which is not due to social interaction, such as a common environment or common individual specific characteristics. To address correlation due to environment, I control for growing conditions and subvillage fixed effects. To investigate the likely correlation of unobserved individual characteristics within networks, I perform a placebo study where I exchange adoption of banana cultivation with the adoption of different cash crops as the outcome variables. If for instance the number of RIPAT banana growers in network captures entrepreneurial spirit, it should also be correlated with the adoption of other lucrative cash crops. But since I find that the number of RIPAT banana growers in the network does not explain much of the variation in the adoption of other crops, the results indicate that the strong correlation between adoption of banana cultivation and the number of banana growing RIPAT farmers in network is not fully driven by correlation in unobservables.

The paper is structured as follows: Section 2 provides an overview of the existing literature, followed by an introduction to the context of this paper in section 3. I proceed to outline the empirical strategy including a description of the data and the applied methods in section 4, and subsequently, I provide a thorough discussion of identification issues in section 5. Results and several robustness checks are presented in section 6, while results from the placebo study are shown in section 7. Finally, section 8 concludes.

2 Literature

The study of adoption of new agricultural technologies is not new in the economics literature and has been studied in a variety of countries and settings since the seminal work by Griliches (1957) studying the adoption of hybrid corn in the United States. The literature is vast as indicated by the review studies Feder, Just, and Zilberman (1985) and Evenson and Westphal (1995), and the more recent handbook chapter of Sunding and Zilberman (2001). Foster and Rosenzweig (2010) provide the latest review of micro studies of adoption of technology with a specific focus on social learning in adoption.

A thorough literature review of the economic research on adoption of technologies is beyond the scope of this paper, but I will provide a short overview of the determinants of adoption of a new technology drawing on both sociological and economic literature. This overview will guide the choice of explanatory variables in the econometric analysis. Next, I will mention some key papers from the literature on social interactions and network with a particular focus on identification of social interaction effects, which will provide a background for discussions of identification in section 5. Then I will focus more narrowly on the literature on adoption of technologies and information networks, first with a focus on theoretical models and finally with a review of the limited empirical literature within this field. As I do not develop a theoretic model for social learning in this paper, I will synthesise the implications of the existing models. The empirical studies will serve as a reference point for econometric analysis.

2.1 Determinants of adoption of a new technology

At first, it is good to have a concept of when an agricultural technology is *new*. It matters little whether the technology is actually newly developed; what is important according to Rogers (1995) is whether it is perceived as new by the individual or the local community where it is introduced.

There can be different and several reasons to adopt a new technology. As listed by Sunding and Zilberman (2001, p.:210) the new technology can be "yield-increasing, cost-reducing, quality-enhancing, risk-reducing, environmental-protection increasing, and shelf-life enhancing" or have several of these qualities, and for modelling purposes it is important to distinguish. In the sociological literature, Rogers (1995, pp.:17-18) provides a useful characterisation of innovations that affect the adoption decision. Here, the individual choice to adopt depends on the *perceived relative advantage* of the technology over existing technologies. The advantage can be economic, but also related to social prestige, convenience and satisfaction, and the perception is naturally affected by discussion with peers. Obviously, the economic advantage is the main focus in the economics literature on the subject (see e.g. Foster and Rosenzweig, 2010). A second factor is the *compatibility* of the innovation with values, norms and past experiences. Third, the *complexity* represents the perception of how difficult the innovation is to understand and use. Fourth, the possibility of experimenting on a small scale, the *trialability,* will ease barriers to adoption and finally, the *observability* of the innovation, how visible it is to others, will stimulate discussion about the innovation and a transfer of information about it.

Within the economic literature, the attention is usually brought to the constraints to adoption of technology that serves to explain the empirically observed rather slow diffusion process relative to theoretical predictions with unconstrained, fully informed and rational agents. Feder, Just, and Zilberman (1985, p.:255) mention a long list: "lack of credit, limited access to information, aversion to risk, inadequate farm size, inadequate incentives associated with farm tenure arrangements, insufficient human capital, absence of equipment to relieve labor shortages [...] chaotic supply of complementary inputs [...] and inappropriate transportation infrastructure". These factors can explain the lack of diffusion of a new technology which is thought to be advantageous. Evenson and Westphal (1995) also provide the useful notions of *tacitness* and *circumstantial sensitivity.* The knowledge about how to efficiently apply a new technology is to some degree tacit, in the sense that it is not embodied in machinery and cannot be codified, which hampers the transfer of new technologies even under the similar circumstances. When circumstances differ, the optimal techniques may differ too, leaving room for adjustment of a new technology to local circumstances. Both concepts leave scope for learning-by-doing and social learning as they hamper one-size-fits-all information about a new technology.

2.2 Social interactions and networks

Social learning generally plays the role of relaxing an information constraint. The information network affects the farmer's perception of the relative advantage of a new technology, along with the perceived complexity. The possibility to try the new technology on a small scale fosters learning-by-doing, while visible yields of the new technology stimulates discussion and exchange of ideas within the information network. Risk averse farmers can be reluctant to try new unknown technologies, but the uncertainty can be mitigated through social learning. To the extent that farmers are aware of the growing circumstances of the members of their information network, they can add more weight to information from farmers in their network with similar growing conditions. In these ways, the information network of the farmer can play an important role for the decision to adopt a new technology. Understanding social learning in the context of a project-introduced new agricultural technology can be important for the future design of development projects.

Among sociologists, social psychologists and anthropologists, social interactions have long been at the centre of the analysis and a review of this literature is way beyond the scope of this paper. Within economics, social learning in developing countries has received increased attention over the past two decades and according to Munshi (2007), the theoretical literature has grown more rapidly than the empirical literature because of the difficulty of identifying social learning effects. This paper will hence add to the thin empirical literature placing identification at the centre of the analysis.

Studies of identification of social interactions constitute a branch of literature per se (e.g. Brock and Durlauf, 2001, 2007; Bramoullé, Djebbari, and Fortin, 2009; Manski, 2000, forthcoming) spawned by the pioneer study of Manski (1993). He addresses the identification of *endogenous social effects* which he defines as the impact of one agent's actions on another agent's actions. In Manski (2000), he describes three different channels through which the endogenous social effects can work: constraints, preferences and expectations. An example of social interactions through constraints is the case of congestion where an agent's use of a road constrains other agents' use of the road. Preference interactions are related to e.g. conformism, where the preferences of a group of agents can affect the preferences of the individual because he wants to be like the others. Finally, agents can affect each others expectations, e.g. expectations about the unknown profitability of a new technology. I will focus on the latter type of social interactions in this paper. Manski (1993) introduces the *reflection problem* that occurs, when a researcher observes the behaviour in a group and tries to infer the effect of the average group behaviour on the individual group members. In a linear model it is difficult to distinguish whether the average group behaviour affects the behaviour of the individual (the endogenous social effect that we are after) or whether group behaviour is simply the aggregation of individual behaviours.

Manski employs the useful vocabulary of endogenous social interaction effects, contextual effects and correlated effects. The *endogenous effects* describe how the behaviour of the individual is affected by the average behaviour in the peer group. The *contextual effects* (or exogenous social effects) cover how the behaviour of the individual is affected by the exogenous characteristics of the group. The *correlated effects* are covariation in behaviour within a group due to similar unobserved individual characteristics or because group members face a similar environment. I will use this terminology to structure my identification strategy as described in section 5. Where Manski (1993) focuses on identification in a linear model, Brock and Durlauf (2001) consider identification in the case of a binary outcome, such as adoption of technology. They both treat social interactions as relations within a defined group where everyone in the group relate to each other and are affected by the expected mean behaviour in the group. Bramoullé, Djebbari, and Fortin (2009) add to this literature by considering *social networks* rather than self-referential groups. The social networks imply that two persons who are connected to a third person are not necessarily connected to each other. He exploits this fact to establish identification in a linear model.

One way to obtain a plausible identification strategy is to have experimental variation in the network variables of interest. Kremer and Miguel (2007) study social learning and deworming behaviour in Kenya and exploit that schools were randomly chosen for the deworming project. When surveying the social links of parents they hence obtain exogenous variation in the number of links to treated schools conditional on the total number of links. Similarly, Duflo and Saez (2002) use experimental data to identify social interactions in retirement plan decisions at a large American university. Unfortunately, project participation is not randomly allocated in my study, which complicates the identification of social network effects. I treat the issue rigorously in section 5.

In the following, I will focus more narrowly on articles on information networks and adoption of technology. Though several articles provide both theoretical and empirical contributions to the literature I will distinguish between the theoretical models and the empirical evidence.

2.3 Modelling information networks and adoption

In a world of perfect information there is no room for social learning as all farmers would know the expected relative yield and the associated risk of a new technology. Subject to other constraints such as credit and input constraints, all farmers would apply the optimal technologies the moment they became accessible. However, when information about new technologies is imperfect, the adoption rate in the farmer's network and hence access to information about the new technology could potentially affect the adoption decision of the farmer.

Bayesian models dominate the theoretical literature on information networks and adoption (e.g. Banerjee, 1992; Besley and Case, 1993; Foster and Rosenzweig, 1995; Munshi, 2004) with O'Mara (1971) being among the first[4] to employ a scientific Bayesian model where farmers update their beliefs about the pay-off of the new crop variety from period to period through the experience with the crop in the surrounding community.

The *target input model* is an example of a Bayesian model that incorporates both learning by doing (Jovanovic and Nyarko, 1994) but also learning from others in the modified versions of Foster and Rosenzweig (1995) and Bardhan and Udry (1999). The form of the new technology is known by the agents in the model, but the optimal amount of input is unknown and stochastic and hence, subject to learning. So the agents know the functional relationship between input and output, but they do not know the mean target input which optimises the yield of the crop. Applied levels of input above or below the target input will result in suboptimal yields, and the new technology is assumed to be superior to existing technologies given that the target input is applied. This implies that profitability of the new technology grows over time as farmers learn about the target input either from own experimentation or experimentation in her/his network. Thus, individual adoption will depend on the adoption rate in the information network of the individual. However, learning from others also leaves room for *strategic delay* in adoption as early

[4]According to Feder, Just, and Zilberman (1985, p.:264).

adopters may forego profits in the learning phase compared to traditional technologies while late adopters can free-ride on the experimentation of the early adopters, as pointed out by Bardhan and Udry (1999).

Where uncertainty about optimal input creates scope for learning in the target input model, other Bayesian models focus on uncertainty of the expected yield or relative profitability of the new technology (Munshi, 2004; Besley and Case, 1993; Banerjee, 1992). In Besley and Case (1993), the planting decision depends on the expected future gains of the new technology relative to the traditional technology, and knowledge is a public good, so farmers update their beliefs about the profitability of the technology either through experimentation on their own farm or among other farmers in the village. Profitability is stochastic but not improving over time due to learning as in the target input model. That is, learning improves the knowledge about the profitability of the new technology, not the profitability itself. This is also the case in the model of Munshi (2004). He models the acreage allocated to the new technology instead of the binary decision to adopt or not. The new technology is more profitable but also more risky and since the farmers are risk averse they may choose to allocate only part of their land to the new technology. For maximising expected utility, he applies the mean-standard deviation approach where the mean yield of the new technology is subject to learning which in turn reduces the uncertainty and hence the standard deviation of the expected utility. Learning takes place through observing acreage decisions and yields of other farmers in the village.

An early model within this strain of literature was that of Banerjee (1992). He models herd behaviour by letting the rational decision maker follow what previous decision makers had done in her place assuming that they had some important information. This leads to an inefficient equilibrium which is very volatile to the information received by the first decision maker. Ellison and Fudenberg (1993) suggest a model where the decision makers base their choices on rules of thumb instead of full Bayesian learning, and show that this model can actually lead to a fairly efficient long-run state in a number of cases even though the individual decision rules are quite naive.

Taken together, the theoretical models generally predict that individual adoption is positively affected by adoption in the information network. This is what I want to test empirically. An exemption is the effect of strategic delay in the target input model which would imply an inverted U-shaped relationship between the number of adopters in the network and the adoption decision. I will also investigate if this is case in my data. But first I will walk through previous empirical findings.

2.4 Empirical literature on adoption and networks

As previously mentioned, the empirical literature on adoption of agricultural technologies and information networks is rather thin, presumably due to the inherent difficulties in identifying social network effects. In order for the researcher to know the relevant reference group, data on individual self-reported networks may be needed which increases data requirements for studies of social networks. Durlauf and Ioannides (2010, p.:470) distinguish between social interactions models where individuals are affected by “crude aggregates”, whereas social networks models focus on the “microstructure of interactions”. Earlier studies such as Besley and Case (1994); Foster and Rosenzweig (1995) and Munshi (2004) use villages as information networks, (i.e. “crude aggregates”), while more recent studies such as Conley and Udry (2010) and Bandiera and Rasul (2006) rely on self-reported social networks. I will focus on the studies with self-reported networks, corresponding to the way I measure information networks.

Bandiera and Rasul (2006) study the decision to adopt sunflower cultivation in Northern Mozambique. They explore adoption within a project using cross-sectional data from 198 households at the year of project implementation. They measure the network by the number of people the farmer knows who grow sunflowers and distinguish between family and neighbours or friends. They find that the number of adopters in the network has an inverted U-shaped relationship with the propensity to adopt as predicted by the target input model and interpret the results as an indication of the presence of strategic delay. There are several obstacles to their identification strategy which they also point out in the paper. They rely heavily on the found inversed U-shaped relationship between number of adopters in network and own adoption decision. They touch upon issues for identification such as correlation in unobservables within networks, endogenous network formation leading to reverse causality, and economies of scale in marketing leading to correlation in choices. However, they argue that these relationships between network and adoption are all monotone (either positive or negative) and cannot explain the inverted U-shaped relationship. They further conclude that an inverted U-shape relationship between unobserved ability and the adoption decision does not seem to drive the results as they are robust to the omission of the most productive cashew nut cultivators.

One of the most convincing empirical evidence of social learning effects in adoption of technology is found in the study of Conley and Udry (2010). They study social learning in the use of fertilizer for pineapple production in Ghana using unbalanced panel data from 15 rounds amounting to 107 changes in fertilizer use by 47 farmers. The information network of the farmer is constructed by drawing a random sample of seven other individuals from the sample and asking if the farmer ever has gone to the person for farming advice. They find that farmers respond to "bad news" in their information network by adjusting fertilizer use, where "bad news" is defined as profits below expectations for a given level of fertilizer. If the farmer receives "bad news" about the same amount of fertilizer that she applied in the previous period, she is more likely to change her fertilizer use. If she receives "bad news" about a different amount of fertilizer than she used previously, she is less likely to change her own use of fertilizer. In particular, novice farmers are more prone to change their fertilizer input. The identification strategy relies on the panel structure of their data, where past outcomes in network affect current input decisions. They control for growing conditions and financial neighbourhoods to ensure that these are not driving the results. Endogenous network formation would not invalidate the identification strategy as they do not consider adoption of pineapple cultivation on the external margin.

Van den Broeck and Dercon (2007) provide a study of social interactions in banana cultivation in Tanzania. They consider adoption of different agricultural techniques within banana cultivation to stem a decline in yields using full census data from one village amounting to 119 households surveyed five times during one year. They both measure networks as kinship-based, insurance network and geographical neighbours. They find that technique adoption is positively correlated with average technique adoption in all three kinds of network, but several caveats apply to this result. The technique adoption in network is measured contemporaneously which exacerbates the problem of omitted variables, such as shocks to banana cultivation in the village and correlation in unobserved individual characteristics within networks. The issue of endogenous network formation is mainly relevant for the insurance network. Furthermore, to distinguish between imitation and learning they consider how lagged average banana yield in network is correlated with current individual banana yield and only find a correlation within kinship networks if the highest productive farmer in network is excluded. Spatial and serial correlation in yields is not considered which hampers identification.

In the light of these empirical studies, I argue that there is scope for more empirical evidence on the role of social networks in adoption of new technologies.

3 Context

I study the adoption of improved banana cultivation among farmers in the rural parts of the Arumeru district in northern Tanzania. It is a dry area where cultivation of bananas using traditional methods is limited because the average precipitation is below what is usually considered as necessary for banana cultivation.[5] In 2006, the Rockwool Initiative for Poverty Alleviation in Tanzania (RIPAT) project introduced a new way of cultivating bananas and new banana varieties that made larger scale banana plantations possible in the semi-arid area under study.[6] I will focus on the adoption of banana cultivation among the farmers that did not participate in the RIPAT project but who live in the villages where RIPAT was implemented. In that sense, I am studying the second generation adoption which allows me to shed light on the diffusion of a new technology beyond the project that introduced it. First, I will give an overview of the components of the RIPAT

[5] The average precipitation in the survey area is 749 mm per year, whereas 1000 mm per year is usually considered as necessary for banana cultivation.

[6] The new banana varieties and cultivation method were imported and tested by Dr. Ali Mbwana at the Selian Agricultural Research Institute in Arusha, Tanzania.

project to give an understanding of the complex and holistic intervention it has been. Next, I will describe in more detail the new banana cultivation technology under study.

3.1 RIPAT

RIPAT is a holistic agricultural and livestock project that aims to alleviate food insecurity and poverty among the participating households. It is designed and implemented by the local NGO RECODA and funded by the Rockwool Foundation. Up to this date, four RIPAT projects have been initiated, but this study only considers the first RIPAT project which took place in Arumeru district in Northern Tanzania from 2006 to 2009. The RIPAT1 has worked as a pilot project for the subsequent three RIPAT projects, where components of the project have been adjusted according to experience from RIPAT1. RIPAT2 and RIPAT3 were started in 2008 in Arumeru and Karatu district, respectively, while RIPAT4 was started in 2009 in Korogwe district, and all three projects are still ongoing. As I am only using data from RIPAT1 in this paper, I will henceforth refer to RIPAT1 simply as RIPAT. The following presentation builds upon Maguzu, Ringo, and Vesterager (forthcoming).

The main idea of RIPAT is to provide a 'basket of option' from which the farmers can pick the components that best fit their soil, water accessibility, availability of household labour and land, preferences and taste. The basket consists of the following:

- Cultivation of improved varieties of banana, with new cultivation techniques
- Crop diversification, including introduction of improved varieties of planting materials such as cassava, sweet potatoes, maize, pulses (soya, pigeon pea, lablab, cowpea) and vegetable production using supplementary irrigation where possible and appropriate
- Conservation Agriculture (minimum tillage using ox-drawn ripper or special hand hoes, erosion control using cover crops, etc.)
- Improved animal breeds (cattle, goat, sheep, pigs, poultry), and training in improved husbandry and veterinary treatment

- Multipurpose trees for fodder, shade, windbreaks, timber, firewood, soil fertility, erosion control and food (fruit trees - avocado, citrus, mango)
- Post-harvesting technologies, i.e. food storage, processing, utilization and marketing

The project follows a Farmer Field School approach with groups of 30-35 farmers cultivating a common plot where the new agricultural techniques are demonstrated. After learning about the new techniques and improved varieties the participating farmers can then choose what they want to adopt on their own farm. Seedlings and seeds are made available free of charge, but pertaining banana seedlings farmers are obliged to pass on thrice as many seedlings to other farmers as they have received from RECODA. In this way, a dissemination of the new banana varieties was incorporated in the project. RECODA insisted to have the obligation to disseminate in the project in order to work against growing donor dependency and expectations of receiving gifts in the area.[7]

The RIPAT villages were selected in the following way: RECODA asked the District Officer to appoint villages that suffered most from poverty and food insecurity. An additional criterion was that the local leadership was co-operative. I focus on farmers who were not members of a RIPAT group but still it is useful to understand the self-selection into the project since the non-RIPAT farmers are not a random sample. Once eight RIPAT villages had been chosen, RECODA held a village meeting for all villagers where the project was introduced, the farmers could sign up if they wanted to join, and two Farmer Field School groups were formed in each village. RECODA required that participants were already dealing with agriculture. They should be poor, but still have at least one acre of land to ensure that they could benefit from the project. An upper limit of five acres was mentioned, but not rigorously abided. Participants should be socially acceptable people who were willing to share with others. RECODA aimed at no less than 50 percent females with the result that some interested men were asked to let their wives be group members instead. Finally, it was ensured that all subvillages where represented in the RIPAT groups in order to facilitate the dissemination of technology and new varieties. In

[7]There are and has been many development projects in the Arumeru district, and according to RECODA they have affected the expectation in the population to receive gifts and support, creating a growing dependency on donor charity.

the end, the village leadership was responsible for the formation of the two groups and in some villages, they have compromised the poverty criterion and allowed influential people to be members of the RIPAT groups.

Keeping the selection into the project in mind, I can conclude that non-RIPAT households fall in three categories: 1) Households that comply with the criteria, but either signed up and were not selected due to a l on group size; are not interested in changing their livelihood; or prefer to wait and see the benefits of the project instead of participating due to high risk aversion or limited household labour; 2) Households that have too much land and/or are too rich to become group members; and 3) Households that own little or no land, and/or social outcasts and drunkards that are not accepted in a RIPAT group. A comparative analysis of the RIPAT and non-RIPAT households is beyond the scope of this paper, but I return to the potential effects of self-selection into RIPAT on the results in section 5.3.

3.2 Improved banana cultivation

This paper focuses on the adoption of improved banana cultivation and hence, it is appropriate to give an introduction to the agricultural technique and the new varieties introduced with RIPAT.[8] First, in order to understand the benefits of banana cultivation, it is necessary to know a little bit about the agricultural calendar in Northern Tanzania. There are in principle two rainy seasons: The *long rains* in March-May which are the most reliable and the *short rains* in October-November which are very erratic and only occur every other year or even less frequently. The most common crops such as maize and beans are annual crops and can hence be planted once (or twice) a year. Most annual crops rely on water at a certain point in time, so not only the amount of water but also the timing of the water is crucial for the crop yield.[9] Banana plants are perennial crops meaning that they do not have to be replanted every year. Only after ten or more years

[8] This representation is based upon interviews with Dominick Ringo, the Head of RECODA, other RECODA staff, banana farmers interviewed during pilot testing of the questionnaire and discussions with Jens Vesterager, Programme Manager at the Rockwool Foundation.

[9] For instance, maize is harvested 90-120 days after planting, but it is only on day 60-80 that the maize cob is produced. If the maize field is subject to water stress in those 20 days, the maize harvest will fail.

can it be beneficial to renew the banana plants as their yields will depreciate hereafter. It takes approximately one year to establish a banana plant and during the first 4-6 months of establishment it requires a regular supply of water. If it is subject to water stress in this period the plant will dry out. Once the banana plant is established it can go longer periods without water and is at this point more drought resistant than most annual crops. It does not fruit if it does not receive water either from rainfall or irrigation but once it becomes moist after a dry period it will start producing fruits again. This also implies that if the farmer has access to an irrigation channel or another source of water for irrigation the banana plant can fruit several times a year - also in the lean period when the storage of annual crops might start to run out.

Not many farmers grew bananas in this dry area before RECODA introduced the new planting technique. As the soil is hard-pan and cannot contain much humidity it calls for preparation before the banana seedling can grow. One digs a hole of one cubic meter and fills it up with a mixture of top soil and manure where one plants no more than three banana seedlings. There should be three meters distance between the holes in the plantation. When the banana plant is growing, one stool will grow several suckers. As long as the banana plant is moist it should be desuckered such that it only has three to four suckers. The suckers can be fed to animals or sold as seedlings for 500 Tanzanian Shillings (0.3 US dollars) per seedling.

Five improved banana varieties were distributed, all plantains, but if they are left to ripe after picking they turn into sweet bananas. The virtues of the improved varieties are that they respond very well to manure and they grow a shorter trunk than the traditional varieties implying that they are less vulnerable to wind.

Ideally, I should have been able to distinguish between whether farmers are using the new technology and/or improved varieties for banana cultivation. Unfortunately, the questionnaire does not allow us to distinguish between technology and variety but simply refer to growing improved bananas. I henceforth assume that if a farmer answers that she grows improved bananas then she has applied the new technology. This assumption is justified by anthropological field work concluding that it is not possible in this area to establish a banana plantation using the traditional cultivation technique. The results in the paper are robust to changing the measure of adoption of improved banana cultivation to only include households that have more than ten banana stools, and hence are most likely to apply the new cultivation method.

The investment costs related to the establishment of an improved banana plantations are the opportunity costs of land and labour and potentially the purchase of improved banana seedlings, if one cannot get hold of them free of charge from RIPAT farmers or groups. In my sample, 78 percent of the adopting farmers did not pay for their first banana seedling(s). The labour investment related to the establishment of the plantation is large as it is a very strenuous task to dig the big holes in the hard soil by hand and some farmers may even choose to hire casual labour to dig the holes at a rate of around 2,000 Tanzanian Shillings (1.25 US dollars) per hole.[10] Establishment of a larger plantation may hence be costly either in monetary terms or measured in opportunity costs. However, planting one or two banana plants is manageable and affordable for most farmers and a gradual expansion of the banana plantation can then be decided upon after testing the banana plantation on a small scale.

To sum up, I relate improved banana cultivation to the five characteristics of a new technology that affects adoption according to Rogers (1995, pp.:17-18). The demonstration plot of the RIPAT group is typically centrally located and yields are very visible both at the demonstration plot and the individual farms. The small banana plantations are particularly conspicuous in the dry areas where they appear as small oases of green in a dusty light brown landscape. There are no specific barriers to small scale experimentation, i.e. the new technology has a high degree of trialability. The instructions for improved banana cultivation consists of a few rules of thumb and do not require special skills to apply. Hence, the complexity of the technology should not be a high barrier to adoption. With respect to the compatibility of banana cultivation, Gausset (forthcoming) found that bananas and plantains are highly valued for both cooking and brewing in the project area, and they are even a compulsory component of dowry payments among Meru

[10] As noted by the anthropologist, Quentin Gausset, during field studies in the RIPAT villages.

and Arush people who constitute the vast majority of the population in the area. Thus, banana cultivation is very compatible with the values and norms in the community. On these grounds, I will study the *perceived relative advantage* as affected by the information network of the farmer in the remainder of the paper.

4 Empirical analysis

4.1 Data collection

The data collection took place in January 2011 and was funded by the Rockwool Foundation. As a consultant, I developed survey instruments, designed the survey and planned the execution of the data collection and data entry. A thorough description of the data collection can be found in Appendix A.

The main objective of the data collection was to perform an impact evaluation of RIPAT, and for this purpose all RIPAT households were interviewed together with a corresponding number of control households from other villages, which adds up to 2,042 households. In all 38 villages, the local leaders were interviewed about village characteristics and in the 16 RIPAT villages two people from each RIPAT group was also interviewed about the features of the RIPAT group. In addition, 597 non-participating households in RIPAT villages were interviewed with the intent to study the dissemination of banana cultivation. Data on RIPAT and control households were collected both from the RIPAT 1 area in Arumeru District and in the RIPAT 3 area in Karatu District, both situated in Northern Tanzania. However, the data from non-participating households, which is used for this study, were only collected in RIPAT 1 villages as the dissemination of banana cultivation was still very sparse in the RIPAT 3 villages at the time of the data collection preparations. For more details on the data collection among RIPAT and control households, see Appendix A.

Random samples of 100 non-participating households from each RIPAT village were drawn and village leaders were asked about the adoption status of these households.

According to this information, the average adoption rate in the villages is 13 percent. However, village adoption rates vary greatly from only three up to a maximum of 23 percent in one village. Because rather few households had adopted improved banana cultivation in some villages, I would have ended up with a very small sample of adopting households had I simply drawn a random sample among the non-participating households. Thus, I aimed at collecting adopting households and non-adopting households in equal numbers to ensure that the final sample would include enough adopting households which are the households of interest for this study. This corresponds to a case-control study which is widely used within biomedical literature. I return to the implications for the econometric analysis of this sampling strategy in section 4.4.1. Though the aim was to collect adopting and non-adopting households in equal numbers, the sample consists of 224 adopting households and 373 non-adopting households. The main reason for this rather low number of adopting households is that when I constructed the sample I identified households to be growing improved bananas through either the local leadership or through records held by RECODA. It turned out that 73 households that had been identified as adopting did not grow improved bananas when visited during the data collection, while the reverse was only true for 48 households. For more details, see Appendix A.

As the degree of adoption varied considerably across villages, I decided not to have the same sample size of non-RIPAT households in each village. I adjusted the sample according to how many adopting households I could identify in each village such that I oversampled villages with a high village adoption rate.

Among the non-participating households in the RIPAT villages, the main respondent to the questionnaire was either the person who took the decision to grow improved bananas or the person who takes most farming decisions, depending on whether the household had adopted improved bananas or not. This person was interviewed about his/her personal characteristics and information network, and about the members of the household, their farm, crops, livestock, and assets. In addition, the adult female in the household was interviewed about household facilities and food security, and we took anthropometric measures of the children in the household under six years of age.

My main estimation sample is constructed as follows: 597 non-RIPAT households in eight RIPAT villages were interviewed in total. Out of these, 36 households have missing data on information network while 15 households have missing data on other explanatory variables, leaving 546 observations in the sample. The data is not systematically missing from either adopting or non-adopting households. Eleven of the adopting households have planted their first improved bananas before 2006, the year RIPAT was implemented. I interpret this as an indication that they have not planted bananas using the new technique introduced by RECODA and hence they should not be considered as adopting farmers. I omit these observations from the main analysis, leaving 535 households.[11] Furthermore, 15 households moved to the village later than 2006 when RIPAT was implemented. These households have not been exposed to the same information as households who lived in the village throughout the period, so I choose to drop these households from the analysis.[12] This leaves a final sample of 520 households among which 194 are growing improved bananas.

4.2 Information networks

When constructing a network measure one of the first challenges is to specify the boundary of the network under study. Scott and Carrington (2011) summarises three different approaches: A position-based approach where only agents with certain positions are considered; an event-based approach where agents who had participated in a particular event are presumed to constitute the relevant population; and a relation-based approach where the analysis starts with a number of individuals of interest and is expanded by agents related to these individuals in a specified manner. The latter approach is particularly applied in the study of *egocentric networks* where the networks of certain individuals (egos) are studied, as opposed to *whole networks* that focus on all agents within the boundary (Scott and Carrington, 2011). I choose to measure egocentric networks for two reasons: Firstly, this is the same type of network measure applied in Bandiera and Rasul (2006) and Conley and Udry (2010) which are the two empirical papers with direct network mea-

[11]Including them in the sample as either adopting or non-adopting does not alter the results.
[12]Including immigrants in the sample does not alter the results.

sures that come closest to my set-up. Secondly, collecting data on egocentric networks as opposed to whole network is much less time consuming and hence, less costly. The measures of the information network are based on the following three recall questions:

> Think about your relatives and friends and other people that you know. Before you decided to start growing improved bananas, how many people did you discuss farming issues with?
>
> Among these, how many of them were growing improved bananas before you decided to grow improved bananas?
>
> If any of these are RIPAT farmers, could you please give me their names?

These questions are inspired by the network questions in Bandiera and Rasul (2006) and Conley and Udry (2010). However, none of them ask for the network size, but I find it important to control for the size of the network to make sure that a correlation between the network measure and adoption does not occur simply because adopters have larger networks and hence are more likely to know other banana growers.

The questions are sequential in the sense that the number of farmers mentioned will be a subset of the number of farmers mentioned in the previous question. This implies that only RIPAT farmers who grow improved bananas are listed. Henceforth, when I mention RIPAT farmers in the information network they will implicitly belong to the subset of RIPAT farmers who grow improved bananas.

The wording above is the questions posed to the adopting farmers. The questions to the non-adopting farmers do not refer to the time "before you decided to start growing improved bananas" and are phrased in present tense. That is, the network measure does not refer to the same time for adopting and non-adopting farmers. I argue that this would cause a downward bias (if any) in the estimate of the effect of network on adoption since I expect a positive trend in the number of banana growers in a given network over the period. Conditional on the network size, a positive trend would cause non-adopting farmers to have relatively more banana growers in their network than adopting farmers, ceteris paribus, as their network measure refers to a later point in time.

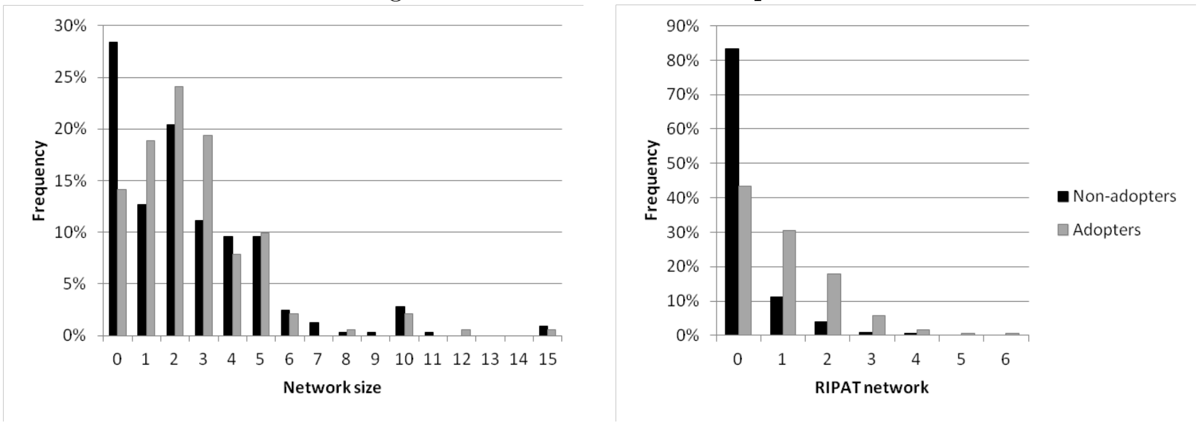
There are several reasons for the specific time reference applied in the network questions. First of all, if I had asked for the current network it is most likely that the network measure would have been endogenous. Obviously, adopting farmers could start discussing farming issues with other banana growers once they had adopted. Had I instead asked a recall question about the network prior to project implementation, no one in the network would be growing improved bananas. I could have asked about the adoption history of this prior network, but I considered that too complicated and time consuming. The chosen phrasing has the virtue of being simple and easy to understand and remember. One drawback could be if farmers who were interested in improved banana cultivation, started discussing with other banana growers before taking the decision to adopt *and* mentioned these farmers when asked the above questions. I return to this issue in section 5.1.

Other applications exploit variation in the strength of the tie in the network referring back to the seminal work of Granovetter (1973) who distinguish between *strong* and *weak ties* as measured for instance by the amount of time spent in the relationship. Kremer and Miguel (2007, p.:1025) measure social links, e.g. as "the five friends they speak with most frequently" to capture the closest social links, while Bandiera and Rasul (2006) distinguish between family and friends in the network. I have not collected data on the type or strength of the tie due to tight budget constraints in the data collection. I could calculate the geographical distances to RIPAT farmers in network using Global Positioning System (GPS) data, but that measure of the strength of the tie would be correlated with the growing conditions that the farmers face (see section 5.3).

Turning to the data, Figure 4.1 shows the network size and the number of RIPAT farmers in network for adopting and non-adopting farmers, respectively. The distribution of network size presented in Figure 4.1A is not too different across adopting and non-adopting farmers, though a larger share of non-adopting farmers do not discuss farming issues with anyone, i.e. have a network size equal to zero. In addition, Table 4.1 shows that the mean network size is not significantly different in the two groups.[13] On the other

[13]This is robust to exclusion of the five households with 20 or more people in their network. However, if I only consider farmers with less than six farmers in their network, which is 92 percent of the sample, adopting farmers have a significantly larger network than non-adopters.

Figure 4.1: Network and adoption



A. *Number of farmers in network* B. *Number of RIPAT farmers in network*

Note: Data from five households with 20 or more farmers in their network are excluded in the two figures.

hand, Figure 4.1B provides a clear picture: adopting farmers have more RIPAT farmers in their information network than non-adopting farmers, which is also confirmed in Table 4.1. The t-test of equal means is rejected at any relevant significance level. The table also shows that adopting farmers discuss farming issues with more non-RIPAT banana growers than non-adopting farmers, a difference that is just significant at the five percent level.

4.3 Other observable characteristics

Table 4.1 summarises the observable farmer and household characteristics that are relevant for adoption of improved banana cultivation, divided by adoption status. The last column gives the p-value from a t-test of equal means for adopting and non-adopting households.

I distinguish between farmer and household level characteristics. At the farmer level, I control for the usual suspects of gender, age, literacy, numeracy and religion. In addition, I follow Bandiera and Rasul (2006) and control for whether the farmer participates in an NGO project (other than RIPAT) or has participated in the past. Table 4.1 shows that adopting and non-adopting farmers are equally likely to be female farmers and there is not a significant age differential between the two groups. They are also equally likely to be literate and have a certain level of numeracy. The farmer's reading and math skills shows to be unimportant for adoption and is hence only included as a robustness check and not in the main analysis. Catholic farmers are significantly more likely to adopt

Table 4.1: Summary statistics

Network variables	Full sample		Adopting		Non-adopting		P-value
No. of RIPAT banana growers	0.51	(0.90)	0.95	(1.09)	0.25	(0.63)	0.000
No. of non-RIPAT banana growers	0.32	(1.86)	0.54	(2.73)	0.19	(1.03)	0.040
Network size	2.81	(4.08)	2.95	(3.78)	2.73	(4.26)	0.541
Farmer characteristics	Full sample		Adopting		Non-adopting		P-value
Farmer is female	0.22	(0.41)	0.23	(0.42)	0.21	(0.41)	0.748
Age of farmer	44.78	(15.61)	43.93	(13.28)	45.29	(16.85)	0.340
Farmer can read*	0.20	(0.40)	0.18	(0.38)	0.22	(0.41)	0.243
Farmer is good at math*	0.39	(0.49)	0.41	(0.49)	0.38	(0.49)	0.429
Farmer is Catholic	0.08	(0.28)	0.11	(0.32)	0.06	(0.25)	0.050
Farmer is Muslim	0.05	(0.23)	0.03	(0.16)	0.07	(0.26)	0.029
Farmer has other religion	0.25	(0.43)	0.25	(0.44)	0.24	(0.43)	0.794
Participates in NGO project	0.24	(0.43)	0.27	(0.45)	0.21	(0.41)	0.130
Household characteristics	Full sample		Adopting		Non-adopting		P-value
Highest education level, years	7.78	(2.24)	8.11	(2.11)	7.58	(2.30)	0.009
Household labour in persons	2.53	(1.50)	2.83	(1.61)	2.35	(1.40)	0.000
Head of household is widow(er)	0.09	(0.29)	0.06	(0.23)	0.11	(0.32)	0.030
Wealth (poverty score)	44.13	(15.05)	46.19	(14.70)	42.90	(15.15)	0.016
Owns mobile phone*	0.61	(0.49)	0.63	(0.48)	0.60	(0.49)	0.533
Number of acres in 2006	4.18	(5.99)	4.10	(4.35)	4.23	(6.78)	0.819
Has grown traditional bananas	0.36	(0.48)	0.46	(0.50)	0.30	(0.46)	0.000
Number of crops grown in 2010	3.94	(1.91)	4.43	(1.93)	3.65	(1.85)	0.000
Grows improved maize in 2009*	0.52	(0.50)	0.66	(0.48)	0.44	(0.50)	0.000
Grows traditional maize in 2009*	0.66	(0.47)	0.56	(0.50)	0.72	(0.45)	0.000
Grows beans in 2009*	0.79	(0.40)	0.86	(0.35)	0.76	(0.43)	0.007
Grows vegetables in 2009*	0.43	(0.50)	0.55	(0.50)	0.36	(0.48)	0.000
No. banana growers within 0.5km	10.57	(8.40)	13.40	(8.63)	8.89	(7.80)	0.000
Use irrigation channel*	0.91	(0.29)	0.96	(0.20)	0.88	(0.33)	0.002
Average yearly rainfall in mm*	749.31	(50.54)	746.90	(39.05)	750.78	(56.40)	0.402
Dist. to nearest waterway, km*	1.17	(0.88)	1.13	(0.75)	1.19	(0.95)	0.400
Placebo outcomes	Full sample		Adopting		Non-adopting		P-value
Grows vegetables in 2010	0.50	(0.50)	0.60	(0.49)	0.44	(0.50)	0.000
Grows sunflower in 2010	0.13	(0.34)	0.15	(0.36)	0.12	(0.32)	0.279
Grows sugarcane in 2010	0.08	(0.27)	0.12	(0.32)	0.05	(0.22)	0.006
Observations	520		194		326		

Notes: Sample means, standard deviations in parantheses. Last column: P-value from t-test of equal means across adoption. Farmer and household characteristics marked with an asteriks (*) are not included in the baseline specification.

compared to Protestants on average, whereas Muslim farmers are significantly less likely to adopt. There is no difference in the propensity to adopt among other religions which is a combined group of both traditional religion practitioners, Seventh Day Adventists and other groups that do not fall into the three main religion groups. Bandiera and Rasul (2006) find that a larger share of adopting farmers participate or have participated in other NGO projects compared to non-adopters, but there is not a significant difference in project participation across adoption status in these data.

At the household level I consider different components of the household structure, namely the highest education level obtained, available household labour, whether the household head is a widow(er), the wealth of the household, and the farm size. This range of variables address constraints to adoption with respect to inputs to agricultural production: capital, labour, human capital, and land. In addition, I control for household ownership of a mobile phone which may capture access to information other than through information network. I further include variables that capture agricultural practices and growing conditions that may correlate with both network and adoption.

Adopting households have achieved significantly higher education levels than non-adopting households. They also have significantly more household labour accessible which is measured by the number of household members who can do hard manual labour to full extent. There are significantly fewer widow(er)s among adopting households and they are also significantly more wealthy, where wealth is measured by a Tanzanian poverty score developed by Schreiner (2011). Ownership of a mobile phone could capture access to information and affect the information network, but there is no significant difference across adoption status. Neither is their a difference in the farm size of adopting and non-adopting households as measured in 2006. I choose to include acreage in 2006 instead of in 2011 due to the potential endogeneity of the farm size: if banana cultivation is profitable then early adopters may have expanded their farm.[14] Different agricultural practices are highly correlated with adoption. Households that grow or have grown traditional bananas and households that grow a higher number of different crops (net of traditional and improved bananas) are both more likely to adopt improved banana cultivation. The number of different crops grown can both capture preferences for diversification of risk and an entrepreneurial spirit among household members. To check if adoption only takes place within groups of farmers who grew another crop prior to adoption, I control for cultivation of improved and traditional maize, beans and vegetables in 2009. Adopting households are significantly more likely to grow improved maize, beans and vegetables and less likely to grow traditional maize than non-adopting households. The number of banana growers

[14] As 97.5 percent of the sample owns at least some of their land and 83.9 percent owns all of the crop land they cultivate, inadequate incentives with respect to farm tenure arrangements should not be a constraint. Hence, I do not distinguish between whether the household owns or rent in the land that they cultivate.

within a radius of 0.5 kilometres from the household is significantly higher for adopting than non-adopting households. It is captured by the number of banana growers in our sample within 0.5 kilometres from the household where distance is measured as distance between GPS points taken at the household's compound. As we have not collected census data the measure is not complete, but it is a good proxy for the growing conditions that the household faces.[15] The use of irrigation channels is more common among adopting households, but there is no significant difference between adopting and non-adopting households in historical rainfall or distance to the nearest waterway. [16]

The last panel in Table 4.1 shows the means and standard deviations of the outcomes variables used for the placebo study. Half of the sample grows vegetables while only 13 percent grows sunflowers and eight percent grows sugarcane in the sample. It appears that vegetable and sugarcane cultivation is positively correlated with improved banana cultivation while sunflower cultivation is not.

4.4 Methodology

I analyse the impact of the number of banana growers in a farmer's network on the farmer's propensity to adopt, using a latent variable approach. I choose not to distinguish between the farmer and household level and use the index i for both levels, as I only have data on one farmer in each household. I do not explicitly model through which channels the information network affects the *perceived relative profitability of banana cultivation*, $a^*_{i,t}$, but I lean on the implications of the theoretical models summarised in section 2.3. Whenever the farmer perceives banana cultivation to be more profitable than alternative crops ($a^*_{i,t} > 0$), the farmer will choose to adopt, where the observed binary adoption choice is denoted $a_{i,t}$:

[15] Once the number of banana growers within a 0.5km radius is controlled for, use of irrigation channel, historical rainfall at the household level and distance to nearest waterway becomes insignificant.

[16] Historical rainfall data is compiled by Hijmans, Cameron, Parra, Jones, and Jarvis (2005). They have interpolated data on precipitation from the period 1950-2000 down to a resolution of 1x1 km using the thin-plate smoothing spline algorithm implemented in the ANUSPLIN package for interpolation, using latitude, longitude, and elevation as independent variables. The data is available from http://www.worldclim.org/. Data on waterways is available at http://download.geofabrik.de/osm/africa/, and distance from household to the nearest waterway is calculated using the 'geodist' package in Stata.

$$a_{i,t}^* = f(N_{i,t-1}, R_{i,t-1}, B_{i,t-1}, \bar{X}_{-i}, X_i, G_i, \alpha_s, \eta_i)$$

$$a_{i,t} = \mathbf{1}(a_{i,t}^* > 0)$$

The perceived relative profitability of banana cultivation is affected by the network through different channels: The size of the network, $N_{i,t-1}$, and the numbers of banana growing RIPAT farmers and non-RIPAT farmers in the network, $R_{i,t-1}$ and $B_{i,t-1}$, respectively. All three measures are indexed with $t-1$ to indicate that they refer to the time prior to the adoption decision for the adopting farmers. Generally, the models summarised in section 2.3 imply a positive relationship between the number of adopters in the network and the propensity to adopt, at least for low numbers of adopters in the network. For higher levels of adoption in network, the target input model implies a decreasing marginal effect due to strategic delay. These effects are denoted *endogenous effects* in the terminology employed by Manski (1993). Furthermore, perceived relative profitability can be affected by networks through *contextual effects,* which refer to the impact of the average farmer and household characteristics in the network, here captured by $\bar{X}_{-i}$. In applications where networks are measured by a position- or event-based approach, (e.g. studies of peer effects in a classroom), it is common to have data on $\bar{X}_{-i}$, (e.g. average family income among class members), but in studies of ego-centric networks it is more rare and to my knowledge, Van den Broeck and Dercon (2007) is the only study within adoption of agricultural technologies in developing countries, that has data to on the characteristics of the network members. As I only have data on characteristics of the RIPAT farmers in the network, I cannot control for $\bar{X}_{-i}$, which is why it is excluded from the empirical specification below. I return to the matter in section 5.2.

In addition, the farmers perception of the relative profitability of banana cultivation is affected by farmer and household characteristics, X_i, growing conditions, G_i, subvillage characteristics, α_s,[17] and individual unobserved characteristics, η_i, where s denotes the

[17] A village is divided into two to five subvillages, each constituting a political unit with a local subvillage leader. There are 24 subvillages in the data. In some villages, the subvillages are contiguous, but in other

subvillage of farmer i. Though these four groups of characteristics may vary over time, they are not indexed by t as they are assumed constant in the period from introduction of RIPAT in 2006 until data collection 2011.

In order to estimate the partial effects I assume that the factors affect the perceived profitability linearly:

$$a^*_{i,t} = \beta_1 R_{i,t-1} + \beta_2 B_{i,t-1} + \beta_3 N_{i,t-1} + \delta X_i + \gamma G_i + \alpha_s + \eta_i$$

I will relax the assumption of linearity of the network variables to investigate the role of strategic delay in section 6.2. Assuming that η is logistically distributed, I can write the propensity to adopt as

$$P(a_{i,t} = 1) = P(\eta_i > -(\theta Z_i + \alpha_s)) = \Lambda(\theta Z_i + \alpha_s) = \frac{exp(\theta Z_i + \alpha_s)}{1 + exp(\theta Z_i + \alpha_s)},$$

where $\theta Z_i = \beta_1 R_{i,t-1} + \beta_2 B_{i,t-1} + \beta_3 N_{i,t-1} + \delta X_i + \gamma G_i$ and α_s is treated as fixed effects. I can then retrieve the logit coefficients by Maximum Likelihood Estimation.

4.4.1 The logit model

There are two reasons to apply the logit model. The primary reason is that case-control data only results in inconsistency of the constant term estimated in the logit model while the logit parameters of the covariates are consistently estimated. Secondarily, the logit model is convenient for handling fixed effects with a binary outcome variable.

Case-control data is common within epidemiology, because one would need enormous random samples in order to study diseases that are not very frequent in the population. The seminal paper in the biomedical literature on the subject is written by Prentice and Pyke (1979) who show that logistic regression coefficients and odds ratios are estimated consistently by applying the original logistic regression model to case-control data. Maddala (1991, p.:793) argues that the results from Prentice and Pyke (1979) cannot

villages there is a larger distance in between subvillages and soil quality may differ across subvillages within the same village. Hence, I control for subvillage fixed effects instead of village fixed effects.

be generalised to the probit estimator or the linear probability model. The sampling method of case-control data is referred to as choice-based sampling within the economics literature, as the sampling method is relevant when considering discrete choices that are rare in the population. Within this strand of literature, McFadden (1973) is the first to show the virtues of the logistic regression with choice-based sampling as acknowledged by Manski and Lerman (1977). Appendix B derives the consistency of the logistic regression coefficients in the case of choice-based sampling and subvillage fixed effects.

The fixed effects methods applied to linear models are not straight-forward to generalise to non-linear models, because they exactly exploit linearity to make unobserved elements cancel out. Due to the functional form of the logit model, it is possible to find the conditional joint distribution of the outcomes in a cluster that does not depend on the unobserved cluster fixed effect Chamberlain (1982).[18] Only observations in clusters with variation in the outcome contribute to the estimation of the parameters and the log likelihood function will be the sum of the conditional log likelihoods for each cluster.

Marginal effects cannot be computed when accounting for fixed effects, but it is actually possible to calculate the marginal effects from estimates based on case-control data (without fixed effects) by correcting for the propensity to adopt in the population. Even though the data allows me to estimate the adoption propensity in the population, I refrain from such calculations and restrict the analysis to consider sign, significance and relative magnitude of the logistic regression coefficients.[19]

[18] The conditional logit model assumes that the explanatory variables are strictly exogenous conditional on the fixed effect, and that the outcomes are independent conditional on the explanatory variables and the fixed effect.

[19] For marginal changes the relative magnitude of the logistic coefficients corresponds to the relative marginal effects. Proof: Let $\Omega = \beta_1 R_{i,t-1} + \beta_2 B_{i,t-1} + \beta_3 N_{i,t-1} + \delta X_i + \gamma G_i + \alpha_s$. The first derivative of the logistic distribution function with respect to R_i is $\partial\Lambda(\Omega)/\partial R_i = \left[\beta_1 exp(\Omega)(1+exp(\Omega)) - \beta_1 exp(\Omega)^2\right]/[1+exp(\Omega)]^2 = \beta_1 exp(\Omega)/[1+exp(\Omega)]^2$ and corresponding for B_i. Hence, $[\partial\Lambda(\Omega)/\partial R_i]/[\partial\Lambda(\Omega)/\partial B_i] = \beta_1/\beta_2$. For discrete variables, this will only be an approximation.

5 Identification

Identification of social learning effects is inherently difficult because most networks are endogenously shaped through individual choices. Exogenous variation in the network can facilitate identification, but that requires either a randomised intervention or a natural experiment, which is not present in my data. In this section I will go through all the different causes of correlated behaviour which cannot be assigned to social learning and describe how I address them in my econometric analysis.

I consider an ego-centric network that is captured *prior* to the adoption decision by a recall question which makes it less likely to have *reverse causality* from adoption to the network measure. If the network was measured contemporaneously then a correlation between banana growers in network and adoption behaviour could simply be due to the fact that once a farmer has adopted she would start discussing farming issues with other banana growers. I further address the issue of reverse causality in section 5.1. I want to distinguish endogenous social effects from exogenous social effects, where the former describes the effect of *behaviour* in network on the adoption decision of the individual. The latter describes the effect of exogenous characteristics such as education and wealth on the individual adoption behaviour.[20] Following Manski (1993, 2000), these are termed *contextual effects* and they are treated in section 5.2. Endogenous network formation naturally leads to a correlation in individual unobserved characteristics if similar people prefer to share information. In addition, behaviour may be spatially correlated due to growing conditions or institutions. Manski calls these confounding factors *correlated effects* and I address them in section 5.3.

When studying the impact of the information network on adoption behaviour, I cannot distinguish imitative behaviour from learning as pointed out by Foster and Rosenzweig (1995). In order to distinguish between imitation and learning I would need data on productivity which I do not have. Nevertheless, if I assume that farmers are rational, adoption must indicate that they perceive improved banana cultivation to be relatively

[20] Manski (1993) uses the word “exogenous”, however, it should be noted that if networks are endogenously formed then the characteristics of network members may be endogenous too.

advantageous either with respect to profits, household food security or social factors such as prestige.

5.1 Reverse causality

The way I measure information networks makes it important to consider whether these networks are endogenous to the adoption decision. A farmer who is initially interested in banana cultivation (has a high η_i) could start discussing farming issues with other banana farmers before making the final decision. If she reports these people when asked during the survey, then my network measure would be endogenous, $\eta_i \rightarrow R_{i,t-1}, B_{i,t-1}$. Another problem could be if the adopting farmers do not recall correctly who they discussed farming issues with before they decided to adopt. Thus, my network measure could also be endogenous if they mistakenly mention people they befriended after adoption.

A way to test for this is to exploit the existence of *super farmers* in the project villages. These are farmers that have been especially trained by RECODA to teach other farmers how to grow bananas. They are selected by the groups themselves among the best farmers to practice and teach farmers the new methods. The super farmers received special training in June-July 2007, and they were expected to serve both their own groups and the wider community as paraprofessional consultants. Two of the eight villages do not have a super farmer and one of the villages actually has two super farmers. In the following, I disregard the two villages without a super farmer.

These super farmers would be the natural person to turn to for questions about improved banana cultivation before you decide to adopt banana cultivation on your own farm. Hence, they provide an obvious opportunity for testing the described endogeneity. If the reverse causality from adoption to network is present, it should be that the super farmers are better represented in the information networks of the adopting non-RIPAT farmers than regular RIPAT farmers that grow bananas. I test this implication in a dyadic framework.

Table 5.1: Dyads and links

	Link		
Dyad between adopting farmer and ...	No	Yes	Total
... super farmer	924	14	938
... regular RIPAT farmer	6,204	85	6,289
Total	7,128	99	7,227

5.1.1 Dyads

I consider all possible dyads which are pairs of RIPAT farmers (including super farmers) and adopting non-RIPAT farmers within a village.[21] I only consider adopting farmers as the described endogeneity is mainly relevant for this group. I disregard 25 adopting farmers who adopted before the third quarter of 2007, because at this point in time super farmers were not trained yet and hence, the test would not apply to the information network of these farmers. This results in a data set with 7,227 dyads where 938 of them include a super farmer. A link exists if the adopting farmer discusses farming issues with the RIPAT farmer, which is the case in 99 dyads and 14 of these involve a super farmer. I test whether the proportion of links is larger given that the dyad includes a super farmer than given that the dyad includes a regular RIPAT farmer.

5.1.2 Test of equal proportions

Let p_s denote the probability of a link in a dyad given that the RIPAT farmer is super farmer and let p_r denote the probability of a link in a dyad given that it includes a regular RIPAT farmer. I then want to test the following hypothesis,

$$H_0: \; p_r = p_s, \quad H_1: \; p_r < p_s$$

Acceptance of the null hypothesis that the probability of a super farmer link is the same as the probability of a regular RIPAT farmer link would indicate that my network measure is not subject to the described endogeneity. I test against a one-sided alternative

[21] Only two farmers discuss farming issues with RIPAT farmers from another village than their own, and these RIPAT farmers are not super farmers. Hence, restricting the dyads to only include pairs within villages does not alter the results of the analysis.

both because this is the relevant alternative hypothesis, but also because it is the most conservative test as I want to challenge the null hypothesis.

I estimate the probabilities by the sample proportions, e.g. the number of links between adopting farmers and super farmers divided by the total number of dyads between adopting and super farmers. I then test whether they are equal by setting up a test statistic following Berry and Lindgren (1996, p.: 493), assuming that the links are independent,

$$Z = \frac{\hat{p}_s - \hat{p}_r}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{N_s} + \frac{1}{N_r})}}, \qquad \hat{p} = \frac{N_s\hat{p}_s + N_r\hat{p}_r}{N_s + N_r},$$

where $\hat{p}_s$ and $\hat{p}_r$ are the sample proportions of super farmer links and regular RIPAT farmer links, respectively, and N_s and N_r denote the number of dyads with super farmers and regular RIPAT farmers, respectively. Asymptotically, the Z test statistic follows a standard normal distribution.[22]

Table 5.2: Super farmer tests

	$\hat{p}_s$	$\hat{p}_r$	Z	p-value
Full sample	1.49	1.35	0.346	0.365
Kwaugoro	2.49	1.83	0.939	0.174
Marurani	1.16	1.01	0.227	0.410

Table 5.2 shows the proportions along with the value of the test statistic and the p-value for the one-sided alternative hypothesis. The null hypothesis is accepted at any reasonable significance level with a p-value of 0.365 for the full sample. This test assumes homogeneity in the proportions across villages which may not be reasonable to assume. I relax this assumption by considering the villages separately, however, in four of the villages none of the super farmers were known by the adopting farmers prior to the adoption decision. Hence, a formal test is not meaningful for these villages. Table 5.2 includes separate tests for the two villages were links between adopting farmers and super farmers exist, namely Kwaugoro and Marurani. Though the super farmers in Kwaugoro appear to be more known than the regular RIPAT farmers,[23] the proportions are not

[22] Though the sample sizes are relatively large, the normal approximation for Z may be rather inaccurate as the probabilities are quite small. The results could be verified with an exact binomial test.

[23] Kwaugoro is the village with two super farmers.

even significantly different at a ten percent level. The test for Marurani further supports the hypothesis that adopting farmers are not more likely to discuss farming issues with a super farmer than with a regular RIPAT farmer prior to the adoption decision.

Given that the super farmers are selected among the best farmers to practice and teach others, they are likely to be well known in the village prior to being appointed as super farmer. This implies that even in absence of reverse causality I would expect to find that super farmers are better represented in networks of adopting farmers than regular RIPAT farmers. As I find that they are not more likely to be known than a regular RIPAT farmer, this only supports the hypothesis of no reverse causality going from adoption to network formation. It is difficult to think of an argument why the appointed super farmers would be less known in the local communities than other RIPAT farmers.

The test suggests that a reverse causality between adoption and network should not be a major concern for the analysis. However, the test remains indicative as it only applies to network formation with super farmers and not with other RIPAT farmers.

5.2 Contextual effects

To be able to draw conclusions on the presence of social learning it is important to distinguish between the endogenous social interaction effects and the contextual effects. If the farmer's adoption decision is affected by the characteristics of network members regardless of the adoption behaviour of the network members then these characteristics would confound the social learning effect. For instance, if banana growers are on average wealthier than other farmers, then knowing several banana growers implies knowing several wealthy people who may provide informal credit or insurance for you. In that case, a positive correlation between the number of banana growers in network and own adoption is not an evidence for learning but confounded by access to informal credit.

Ideally, I would like to control for the average characteristics of the information network members ($\bar{X}_{-i}$) to ensure that the correlation between adoption and the adoption behaviour in the network is not driven by exogenous characteristics of network members. However, I only have detailed information about the RIPAT farmers in the network and not other network members.[24] To the extent that exogenous characteristics are highly correlated within the network, controlling for farmer characteristics (X) that are expected to affect adoption partly resolves the issue. But it is not sufficient in the case of heterogeneous networks. I do not have data on all network members, but I can exploit the detailed data I have on RIPAT farmers. That allows me to distinguish between knowing wealthier or poorer RIPAT farmers, older or younger, less or more educated, women or men, to see if the effects are heterogeneous. Heterogeneous effects could suggest that the results are driven by contextual effects.

5.3 Correlated effects

I distinguish between correlated effects due to environment (G_i, α_s) or individual unobserved characteristics (η_i).

Farmers within a network may behave similarly because they face the same environment. Agricultural activities may be correlated for neighbouring farmers due to similar growing conditions rather than social network effects. If farmer j is a RIPAT farmer in farmer i's network then correlations in G_i and G_j will cause a spurious correlation between $R_{i,t-1}$ and $a^*_{i,t}$ when G_i is not controlled for. If the village leadership is supporting and promoting banana cultivation in a particular village then a correlation in adoption behaviour within networks in the village would not necessarily indicate the existence of social network effects. If α_s is not controlled for then the correlation between $R_{i,t-1}$ and $a^*_{i,t}$ will be upward biased provided that the RIPAT farmers in farmer i's network are from the same village as farmer i. I address these issues in several ways.

I capture the growing conditions of a farmer by the number of banana growers in my sample within a radius of 0.5 kilometres from the farmer's dwelling as measured by GPS,[25] and I find this measure to be an important determinant of adoption. The GPS measure is taken at the household dwelling and not at the farmer's plot(s), but this should

[24]Bandiera and Rasul (2006) do not have data on the characteristics of network members either. They claim that the U-shaped relationship found between adoption and network cannot be explained by contextual effects as they would be monotone. To my knowledge, the only paper analysing adoption of agricultural technologies and networks with detailed information on network members is that of Van den Broeck and Dercon (2007).

[25]The distance is calculated using the 'geodist' package in Stata.

not add too much noise as the majority of households have plots that are contiguous to their dwelling. As all RIPAT farmers are interviewed and in some villages all identified adopting farmers are interviewed, this measure almost corresponds to the actual number of banana growers within a radius of half a kilometre. However, in the villages where adoption is very wide spread so that the sample does not include all adopting households in the village, it understates the number of adopters within the radius. This is somewhat problematic since it will not capture the full effect of growing conditions in these villages. To mitigate this problem, I could additionally control for the historical rainfall within one square kilometre of the household, the distance from the household to the nearest waterway and whether the household uses an irrigation channel. However, I do not find any of these measures to be important for adoption once the number of neighbouring adopters is controlled for, neither does inclusion of them affect the estimated network effects. I hence consider the measure of adopters within a small radius to be capturing the growing conditions of the farmer. To capture institutional effects I include subvillage fixed effects. The fixed effects also capture general equilibrium effects such as the effect of wide spread adoption in the local market price of bananas.[26]

Another important correlated effect stems from the likely correlation of unobserved individual characteristics within networks which are formed by individual choice. Entrepreneurial farmers may first of all have larger networks and hence, be more likely to know adopting farmers. Thus, I control for network size in all regressions. For a given network size, an entrepreneurial farmer may choose to discuss farming issues with other farmers who are themselves entrepreneurial. Hence, a correlation between their adoption behaviour may simply reflect that they are of the same type rather than being an indication of social learning effects. If eligibility into RIPAT had been randomised, I could have used the random variation in the network of the non-eligible farmers to circumvent this problem (see e.g. Kremer and Miguel, 2007). But because participation in RIPAT was voluntary I must address the potential correlation of unobservables within the network. I do that by performing a placebo study described in section 5.3.1.

[26]However, it should be noted that the majority of farmers face periods of food insecurity and mainly grow bananas for home consumption.

The self-selection into RIPAT creates an additional concern. A farmer who knew many farmers who signed up for RIPAT could have chosen not to sign up simply because she knew that she would learn about the new technologies anyway. Since participation in RIPAT required weekly participation in meetings and joint cultivation of the demonstration plot and hence, many work hours, it is a reasonable concern that some farmers who were initially interested in improved banana cultivation could have chosen not to sign up for RIPAT, because several of their network members had done so. This corresponds to the idea of strategic delay derived from the target input model where a farmer would choose to postpone adoption if she knows sufficiently many adopters allowing her to learn from their experimentation without incurring the cost of experimenting herself. Similarly, a farmer could avoid the opportunity cost of labour related to RIPAT participation if one or more network members had chosen to participate from whom she could get improved banana seedlings and instructions.

If I assume that these farmers would adopt banana cultivation relatively early since they were interested in banana cultivation already at the start of the project, I can split the sample of adopters into early and late adopters to see if there are differential effects. If the network effects only persist among early adopters, they may simply be generated by self-selection mechanisms into RIPAT.

5.3.1 Placebo study

Consider the following situation: Farmer i likes to discuss farming issues with farmer j, because they are both open to new ideas and like to try out new crops on their farm. Farmer j chooses to join RIPAT and establish a banana plantation and while farmer i chooses to plant improved bananas with the help from a super farmer. In this case, no social learning takes place but adoption behaviour of farmer i and farmer j is correlated because of a correlation in η_i and η_j. To examine if the strong correlation I find between adoption behaviour and adopters in network is driven by correlation of unobservables I consider adoption of three other crops: vegetables, sunflowers and sugarcanes which are all profitable cash crops.

Cultivation of vegetables (e.g. onions, tomatoes) is very profitable but also requires access to water and intensive seasonal labour input. Sunflowers can be grown under rather dry conditions and the sunflower oil can be extracted from the seeds with a simple hand press. Sugar cane is a perennial grass that can be grown under varying conditions but access to irrigation water increases yields.[27] If the profitability of banana cultivation dominates the profitability of vegetables, sunflowers and sugarcanes for all farmers then the placebo test has no bite. However, I would argue that this is not the case. Farmers who have access to plenty of water would profit more from vegetables than bananas whereas it might be more beneficial to grow sunflowers for farmers who have very limited access to water. I have chosen these three crops because their profitability relative to banana cultivation varies across farmers conditional on their available inputs. It is not clear that banana cultivation would dominate the cultivation of any of these crops for all farmers.

If the correlation between the number of banana growers in network and adoption of banana cultivation is driven by a correlation of unobservables in the network I would expect the number of banana growers in network to explain variation in the adoption of vegetables, sunflowers and sugarcanes. If knowing more RIPAT farmers is simply a proxy for being more open to new ideas then it should be correlated with the adoption of other crops. Results are presented in section 7.

6 Results and robustness

6.1 Results from baseline specification

Estimated logistic coefficients from the baseline model are shown in Table 6.1. Column (1) presents the simple logistic regression of the propensity to adopt on the three network variables. Discussing farming issues with a banana grower – whether RIPAT or non-RIPAT – is positively correlated with the decision to adopt and the estimated parameters are both significantly different from zero at the five percent level. However, knowing an extra RIPAT farmer appears to be five to six times as important as knowing an extra non-

[27] Information on cultivation of vegetables, sunflowers and sugarcanes is based on conversations with Jens Vesterager, Programme Manager, Rockwool Foundation.

RIPAT farmer who grows improved bananas, given network size, and it is even statistically significant at the one percent level. Network size is not positively correlated with adoption when banana growers in network are controlled for.

In column (2) I include farmer and household characteristics and report conditional logit estimates where subvillage fixed effects are accounted for as described in section 4.4.1. Since these parameters are only identified by variation within subvillages, factors that cause adoption rates to be correlated within subvillages such as soil quality, distance to markets and village institutions are not confounding the network effects. The number of observations is reduced by six households because one subvillage does not have any adopting farmers in the sample.

The estimated parameters for the network variables change less than a standard error from the simple regression. In this specification, knowing a RIPAT banana grower is approximately four times more important for the adoption decision than knowing a non-RIPAT banana grower. The effect is significant at the one percent level while the impact of knowing a non-RIPAT banana grower is only significantly different from zero at the five percent level. So even though I control for a range of farmer and household characteristics and only exploit the variation within subvillages the number of banana growers in the network persists to be strongly correlated with the adoption decision. The network size is now significantly negatively correlated with adoption given the number of banana growers in network which is a bit surprising. I would have expected farmers who are more entrepreneurial to have larger networks and also to be more prone to adopt improved banana cultivation. On the contrary, it appears that discussing farming issues with other people than banana growers makes a farmer less likely to adopt. I will return to this issue in section 6.4.[28]

Turning to farmer characteristics, female farmers are significantly more prone to adopt improved banana cultivation than male farmers and this effect is quite strong approximately corresponding to knowing three non-RIPAT banana growers as compared to none.

[28]If network size is excluded from the model, RIPAT banana growers in network still have a positive and significant effect at the one percent level with a coefficient of 0.857 corresponding to the model in column (2). However, the number of non-RIPAT banana growers in network no longer has a significant impact on adoption. See results in Appendix C.

Table 6.1: Adoption of improved banana cultivation, baseline specification

	(1)	(2)
NETWORK VARIABLES		
RIPAT banana growers in network	1.155***	1.052***
	(0.147)	(0.174)
Non-RIPAT banana growers in network	0.208**	0.251**
	(0.104)	(0.101)
Network size	-0.125	-0.200***
	(0.0759)	(0.0582)
FARMER CHARACTERISTICS		
Farmer is female		0.802**
		(0.341)
Age of farmer		0.0552
		(0.0483)
Age of farmer, squared /100		-0.0617
		(0.0466)
Farmer is Catholic		1.018**
		(0.450)
Farmer is Muslim		-0.135
		(0.605)
Farmer has other religion		-0.154
		(0.306)
Participates in NGO project		0.0256
		(0.291)
HOUSEHOLD CHARACTERISTICS		
Highest education level, years		0.0331
		(0.0674)
Household labor		0.179*
		(0.0990)
Head of household is widow(er)		-0.558
		(0.509)
Wealth (poverty score)		0.0134
		(0.00914)
Number of acres in 2006		0.0774
		(0.0691)
Number of acres in 2006, squared		-0.00377*
		(0.00207)
Grows/has grown traditional bananas		0.505*
		(0.262)
Number of crops grown in 2010		0.226***
		(0.0680)
No. banana growers within 0.5km		0.0361**
		(0.0179)
Subvillage fixed effects	NO	YES
Observations	520	514

Notes: Logit coefficient estimates, constant not reported. Standard errors in parentheses, clustered at subvillage level in column (1). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

This is well in line with anthropologic field work in the area which concludes that women generally have the authority over bananas as compared to beans which is the domain of men (Mogensen and Pedersen, forthcoming). There appears to be an inverse U-shaped relationship between the propensity to adopt and the age of the farmer, though the two terms are not jointly significant at the ten percent level.[29] Though the coefficients are not significant, the pattern is well interpretable. Until the age of 45 there is an increasing relationship between the farmer's age and adoption while the relationship is negative for older farmers. This can be explained by the different phases in a household where a young farmer has to spend time on child rearing, while when the children become older the household can draw on teenage labour force. For older farmers, the children may have left home leaving fewer hands in the family farming activity.

Religion appears to play an important role showing that Catholics are much more likely to adopt than Protestants who constitute the reference group. The correlation corresponds in size to knowing an extra banana growing RIPAT farmer.[30] The other religion dummies are not significantly different from zero. Whether the farmer participates in an NGO project now or has done so in the past does not affect adoption which is opposite to Bandiera and Rasul (2006) who finds project participation to be an important determinant of adoption.

With respect to household characteristics, it is somewhat surprising that education does not appear to have a stronger effect on the decision to adopt, here measured as the highest number of years of schooling anyone has attained in the household.[31] The little importance of education suggests that the new technology is so simple that lack of formal education is not a barrier to adoption. If I had been able to measure the exact techniques applied, I may have been able to demonstrate the role of education in the accuracy of the methods used. On the other hand, household labour appears to have

[29] Wald tests of joint significance: $\chi^2_{(2)} = 2.48$, $p = 0.29$.

[30] Catholics are equally represented among RIPAT and non-RIPAT farmer, so the large coefficient can not be explained by Catholics being reluctant to join RIPAT groups. The role of religion in networks would be an interesting topic for future studies.

[31] As 58 percent of the sample has completed seven years of primary education, I could instead include dummies for having more or less than seven years of education. However, these are not jointly significant either, nor is the education of the farmer included either as years or as dummies.

some impact on adoption though it is only significant at the ten percent level. It is measured as the number of household members who are able to do hard manual labour to a full extent. As the establishment of a banana plantation requires a lot of hard manual labour it is intuitive that the available household labour is positively correlated with adoption. Whether the household head is a widow(er) seems to be negatively correlated with the adoption decision as expected. The estimated coefficient is rather large, but quite inaccurate and hence not statistically significant. Naturally, a widow(er) household has less available household labour, and the strong negative correlation between household labour and the widow dummy explains the large standard errors. The wealth of the household as measured by a poverty score does not appear to be a strong determinant of adoption. Neither does the number of acres the household has access to seem to play a very important role. It is included quadraticly and the two terms are only jointly significant at the ten percent level.[32] Hence, little wealth or limited access to land does not seem to be important barriers to adoption, supporting the trialability of banana cultivation and further suggesting that network effects are not driven by access to credit.

If the household has previously grown traditional bananas, they are more likely to adopt which is intuitive as banana cultivation is exactly compatible with the livelihood of these households. However, it should be noted that this impact only corresponds to half the impact of knowing an extra banana growing RIPAT farmer which stresses the importance of information networks. I include the number of crops the household grew in 2010, net of traditional and improved bananas, to control for the combination of entrepreneurship and preference for risk diversification that would induce the farmer to plant many different crops. The number of crops grown in 2010 is indeed positively correlated with the adoption of banana cultivation and significant at the one percent level. Finally, I control for growing conditions such as soil quality and rainfall by including the number of banana growers within a radius of 0.5 kilometres and the parameter is positive and significantly different from zero at the five percent level.

Among the list of characteristics, the number of RIPAT banana growers in the net-

[32]Wald tests of joint significance: $\chi^2_{(2)} = 5.06$, $p = 0.08$. If number of acres is included linearly the parameter is negative but not significantly different from zero.

work of the farmer prevails as one of the most important determinants of adoption both economically and statistically. The t-statistic of the parameter estimate is 6.25, which is by far the largest t-statistic of the included controls.[33] With respect to the magnitude of the parameter, the effect of the RIPAT network appears to be rather dominant. For instance, it corresponds to the impact of having more than five extra household members who are able to do manual labour to full extent which is quite a large effect taking into account that establishment of a banana plantation requires a lot of labour.

It is noteworthy that the effect of knowing an extra RIPAT banana grower is approximately four times as high as the effect of knowing a non-RIPAT banana grower. This indicates that the diffusion of the improved banana cultivation is strongest from project participants to farmers who discuss farming issues with them, but that the diffusion eases off from the non-participants and onwards. This could be explained by several factors.

Firstly, it could be an indication that the learning effects abate as we move further away from project participants. The RIPAT farmers have been thoroughly trained and are hence convincing teachers while non-RIPAT banana growers affect their network to a smaller extent because they are less experienced and trained than the RIPAT farmers.

Secondly, it takes approximately one year from the establishment of a banana plant before the farmer can harvest seedlings that can be passed on to other farmers in the network. Hence, if non-RIPAT banana growers have planted recently, they may not be able to share seedlings in their network.

Thirdly, 61 percent of the RIPAT farmers who grow bananas and have passed on seedlings to other farmers mention "obligation in the project" as one of the reasons for passing on improved banana seedlings to other farmers. This fact raises the question of whether the adopting farmers simply plant a few banana plants because they received the seedlings as gifts which leads to the high impact of RIPAT network on adoption or whether they really learn the new technology and adopt it because they perceive improved banana cultivation to be advantageous. One way to shed light on this issue is to only consider farmers who asked for the seedlings as opposed to the seedlings being given to them.[34] The

[33] The second largest t-statistic is -3.44 for the parameter of network size.

[34] This can be identified from the following question in the questionnaire: "Were the improved banana

majority went and asked for the seedling (79 percent) and if only these adopting farmers are included in the analysis, the parameter estimate for both the RIPAT and non-RIPAT banana growers in network increases (see estimates in Appendix C). It is not entirely clear whether the differential impact from RIPAT and non-RIPAT banana growers stems from differential learning effects, lag in seedling production or the obligation to share among RIPAT farmers, but this result suggests that the diffusion of improved banana cultivation is not purely driven by the obligation of RIPAT farmers to pass on seedlings to other farmers.

6.2 Functional form

It is not given that network effects are linear and hence, I present estimates for different functional forms of the network variables in Table 6.2. Farmer and household characteristics are included and subvillage fixed effects are accounted for in all specifications.

Column (1) gives the estimated logit coefficients on the external margin: the effect of knowing one or more RIPAT or non-RIPAT banana growers on adoption, respectively. The impact of knowing one or more RIPAT farmers is positive and significant at the one percent level and almost three times larger than the impact of knowing one or more non-RIPAT banana growers. The latter is only significantly different from zero at the ten percent level.

Following Kremer and Miguel (2007), I include the share of RIPAT banana growers and non-RIPAT banana growers in the network instead of the number of people in column (2), still controlling for the network size. The proportion of RIPAT banana growers in the network has a positive impact on the propensity to adopt which is significant at the one percent level. So has the proportion of non-RIPAT banana growers and even though the coefficient is less than two thirds of the coefficient for the RIPAT proportion, a Wald test of equality of the two coefficients yields a p-value of 0.056.[35] Thus, strictly I cannot reject an equal impact of the share of RIPAT and non-RIPAT banana growers in network

seedlings given to you or did you go and ask for the seedlings?"

[35] Wald test of equal coefficients for RIPAT and non-RIPAT banana growers share in network: $\chi^2_{(1)} = 3.64$, $p = 0.056$.

Table 6.2: Adoption of improved banana cultivation, different functional forms of network variables

	(1) External margin	(2) Proportion	(3) Quadratic	(4) Splines
Know RIPAT banana grower(s), (0/1)	1.759***			
	(0.271)			
Know non-RIPAT banana grower(s), (0/1)	0.633*			
	(0.354)			
RIPAT banana grower network share		3.386***		
		(0.468)		
Non-RIPAT banana growers network share		2.031***		
		(0.573)		
RIPAT banana growers			1.733***	
			(0.302)	
RIPAT banana growers, squared			-0.259***	
			(0.0865)	
Non-RIPAT banana growers			0.256	
			(0.209)	
Non-RIPAT banana growers, squared			-0.00245	
			(0.0104)	
1-2 RIPAT banana growers				1.721***
				(0.292)
3+ RIPAT banana growers				2.344***
				(0.697)
1-2 non-RIPAT banana growers				0.548
				(0.375)
3+ non-RIPAT banana growers				0.701
				(0.903)
Network size	-0.0988**	-0.0420	-0.289***	
	(0.0390)	(0.0376)	(0.0757)	
Network size, squared			0.00564***	
			(0.00213)	
Network size: 1-2				0.108
				(0.357)
Network size: 3+				-0.785**
				(0.395)
Observations	514	514	514	514
P-values from Wald tests of splines:	1-2 RIPAT = 3+ RIPAT			0.363
	1-2 non-RIPAT = 3+ non-RIPAT			0.871
	Network size: 1-2 = Network size: 3+			0.002

Notes: Conditional logit coefficient estimates accounting for subvillage fixed effects. Standard errors in parentheses. Farmer and household characteristics are included in all specifications. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

on the propensity to adopt at the five percent significance level.

In column (3) I allow for a quadratic form of the three network variables. The marginal impact of knowing an extra RIPAT banana grower is positive but decreasing until knowing three RIPAT farmers and hereafter the marginal impact of knowing an extra RIPAT farmer is negative. This pattern is significant at the one percent level and corresponds to the inverse U-shape found in Bandiera and Rasul (2006). Nevertheless, I find it important to notice that only 1.4 percent of the farmers in my sample know more than three RIPAT farmers and hence, it would verge on extrapolation to draw conclusion on negative marginal effects in networks with more than three RIPAT farmers. Bandiera and Rasul (2006) conclude that the inverse U-shape they find indicates existence of strategic delay, but after all Bandiera and Rasul (2006) have 16 percent of their sample beyond the peak of the estimated quadratic function. I conclude that the quadratic relationship I find suggest that the impact of knowing a RIPAT banana grower is larger on the external than internal margin. The linear and squared term for non-RIPAT banana growers in network are jointly insignificant,[36] while the estimated coefficients for the network size reveal a U-shaped relationship with the minimum at 24 farmers in network. Only 0.8 percent of the sample has more than 24 farmers in the information network, so basically the relationship between network size and adoption is negative but decreasing (in absolute terms) given the number of banana growers in network.

To further investigate the non-linearity, I consider splines of the network variables: Knowing one to two farmers or knowing three and more farmers, both with respect to RIPAT, non-RIPAT banana growers and farmers in general (network size). I do not find support for negative marginal effects in this specification either as the parameter for knowing three or more RIPAT farmers is not smaller than the parameter for knowing one to two RIPAT farmers. However, as shown in the bottom of Table 6.2 I cannot reject that the impact of knowing three or more is equal to the impact of knowing one or two RIPAT farmers, which provides further support for decreasing marginal returns to knowing an extra RIPAT banana grower. Similar to the quadratic specification, the estimated parameters for knowing non-RIPAT banana growers are jointly insignificant.[37] Taken together, this further supports the result that knowing RIPAT banana growers is more important than knowing non-RIPAT banana growers. With respect to network size, knowing one to two farmers does not affect the adoption decision, while knowing three or more has a negative impact on adoption, ceteris paribus.

Results from the different specifications seem to suggest decreasing marginal returns

[36] Wald test of joint significance: $\chi^2_{(2)} = 3.43$, $p = 0.180$.
[37] Wald test of joint significance: $\chi^2_{(2)} = 2.53$, $p = 0.282$.

to knowing an extra RIPAT farmer. However, for simplicity I will keep the linear baseline specification in the following estimations.

6.3 Contextual effects

I have found a strong relationship between banana growers in network and the adoption decision, but the question is whether this correlation provides evidence for learning effects or is confounded by other factors. If characteristics such as the wealth of the network members affect the adoption decision then a correlation in adoption behaviour within networks may be confounded by access to informal credit. If I had detailed data on all network members, I could control for average characteristics in the network, but as I only have detailed information on the RIPAT farmers in the network, I proceed differently. I split the RIPAT farmers in the network based on four central socioeconomic characteristics: wealth, education, gender, and age, and I examine whether the network effects differ dependent on the characteristics of the network members. No differential effects would support the hypothesis that the network effects found are not driven by contextual effects.

I measure wealth by a poverty score with a range of 0-100 (Schreiner, 2011) and split the sample of RIPAT farmers in networks at the mean poverty score, 47.4.[38] With respect to education, RIPAT farmers are divided into three groups: less than seven years of education (26.9 percent), seven years of education (68.1 percent) and more than seven years of education (5.0 percent of the sample). The gender split is self-explanatory, while the sample is split into young and old farmers at the mean age of 46.9 years.[39]

Table 6.3 shows the estimation results together with tests of equal network effects across different characteristics of network members. Column (1) does not provide support for the hypothesis that the network effect is driven by access to credit as the estimated effect of knowing rich RIPAT banana growers is in fact lower than knowing poor RIPAT banana growers. However, this difference is very far from being significant.

Turning to the split on farmers' education in column (2), it appears that there is a

[38] The median poverty score among the RIPAT farmers in networks is 46.3, and results are robust to splitting the sample at the median.

[39] The median age is 45 which is close to the mean age, and the results are robust to splitting the sample at the median.

Table 6.3: Adoption of improved banana cultivation, split on characteristics of RIPAT network

	(1) Wealth	(2) Education	(3) Gender	(4) Age
Poor RIPAT banana growers	1.047***			
	(0.246)			
Rich RIPAT banana growers	0.950***			
	(0.237)			
RIPAT banana growers with low edu.		0.587		
		(0.364)		
RIPAT banana growers with medium edu.		1.164***		
		(0.216)		
RIPAT banana growers with high edu.		1.009		
		(0.727)		
Male RIPAT banana growers			1.058***	
			(0.204)	
Female RIPAT banana growers			0.829**	
			(0.330)	
Young RIPAT banana growers				1.039***
				(0.255)
Old RIPAT banana growers				1.001***
				(0.260)
Non-RIPAT banana growers in network	0.257**	0.258***	0.254**	0.263***
	(0.1000)	(0.1000)	(0.0989)	(0.0997)
Network size	-0.191***	-0.199***	-0.188***	-0.192***
	(0.0575)	(0.0579)	(0.0576)	(0.0574)
Observations	514	514	514	514
Wald tests of equal effects:[a] χ^2, (p-value)	0.09	1.95	0.36	0.01
	(0.770)	(0.377)	(0.550)	(0.921)

Notes: Conditional logit coefficient estimates accounting for subvillage fixed effects. Standard errors in parentheses. Farmer and household characteristics are included in all specifications. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.
[a] The following is tested: Column (1): Poor = rich (df = 1), Column (2): Low edu. = medium edu. = high edu. (df = 2), Column (3): Male = female (df = 1), Column (4): Young = old (df = 1).

smaller effect from knowing RIPAT banana growers with less than seven years of education, but there is no significant difference on the three estimated network effects for different education categories. The coefficient for knowing RIPAT farmers with a high education is imprecisely estimated as the group is fairly small.[40] To ensure that the acceptance of the null hypothesis is not driven by large standard errors induced by the high education category, I combine the high and medium education category, and again I accept the null of no difference in network effects across education of network members.[41]

[40] Only 13 farmers in the sample know a RIPAT farmer with high education.
[41] Estimation results not reported. Wald test of equal effects for knowing RIPAT farmers with low and medium/high education: $\chi^2_{(1)} = 1.00$, $p = 0.318$.

As can be seen in column (3), the impact of knowing a male RIPAT farmer seems to be larger than knowing a female RIPAT farmer. Nevertheless I again reject that the network effects are differential across gender. Neither do the estimates and test results in column (4) provide evidence for a difference in the network effects across age.

Hence, I conclude that the network effects appear to be rather homogenous across these four socioeconomic characteristics which indicates that the network effects are not driven by contextual effects. At least, the network effects do not seem to be driven by access to informal credit or e.g. by knowing older RIPAT farmers who are maybe more respected and influential in the village. I will address the potential presence of unobserved factors such as the degree of influence in the network that could affect the estimated network effects in the placebo study in section 7.

6.4 Robustness

In Table 6.4 I present different robustness tests of the baseline specification presented in column (3) of Table 6.1. Column (1) includes a range of extra covariates: farmers' literacy and numeracy skills, ownership of mobile phone, use of irrigation channel, historical rainfall and distance to the nearest river or stream. Consequently 12 observations are dropped due to missing values mainly on GPS measures used for determining historical rainfall and distance to nearest waterway. The coefficient for the network variables are all within one standard error of the estimates from the baseline model and a test of joint exclusion of the extra covariates is accepted with a p-value of 0.711.[42]

In column (2) farmers with no people in network or with 20 or more network members are excluded from the sample. Only five farmers discuss farming issues with 20 or more farmers, but 119 farmers (22.9 percent of the sample) do not discuss farming issues with anyone. Figure 4.1 suggested that a larger fraction of non-adopting farmers actually had no information network. To see if the results are driven by unconnected non-adopters I hence also exclude farmers with no information network. The estimate for the number of RIPAT banana growers in network is virtually the same and still significant at the one

[42]Wald test of joint significance of extra covariates: $\chi^2_{(6)} = 3.75$, $p = 0.711$.

Table 6.4: Adoption of improved banana cultivation, robustness of results

	(1) Extra covariates	(2) Excl. outliers	(3) Important crops	(4) Income source	(5) Early adopters	(6) Late adopters
RIPAT banana growers	1.048***	1.088***	0.870***	1.026***	1.412***	0.972***
	(0.178)	(0.189)	(0.230)	(0.177)	(0.310)	(0.185)
Non-RIPAT banana growers	0.242**	0.200	0.164	0.248**	0.397	0.227**
	(0.103)	(0.183)	(0.117)	(0.102)	(0.328)	(0.0993)
Network size	-0.197***	-0.330***	-0.120*	-0.202***	-0.458***	-0.172***
	(0.0595)	(0.0767)	(0.0693)	(0.0592)	(0.165)	(0.0595)
Farmer can read	0.360					
	(0.401)					
Farmer is good at math	0.109					
	(0.275)					
Owns mobile phone	-0.0607					
	(0.286)					
Use irrigation channel	-0.573					
	(0.523)					
Average yearly rainfall in mm	-0.0123					
	(0.00860)					
Dist. to nearest waterway, km	0.234					
	(0.322)					
Grows improved maize, 2009			0.0480			
			(0.415)			
Grows traditional maize, 2009			0.0567			
			(0.404)			
Grows beans, 2009			0.616			
			(0.490)			
Grows vegetables, 2009			-0.303			
			(0.395)			
Other income source				-0.315		
				(0.325)		
Observations	502	391	327	498	263	469

Notes: Conditional logit coefficient estimates accounting for subvillage fixed effects. Standard errors in parentheses. Farmer and household characteristics are included in all specifications. In column (2) farmers with no people in network or with 20 or more network members are excluded from the sample. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

percent level. With respect to non-RIPAT banana growers, the effect is still within one standard error of the baseline result, however no longer significantly different from zero. The negative relationship between adoption and network size is increased in magnitude which is not surprising as the farmers with no information network were mainly non-adopters.

A concern could be that the formation of information networks is affected by previously grown crops which would also be correlated with the adoption of banana cultivation if the profitability of bananas only dominates certain crops. In the baseline specification I control for previous cultivation of traditional banana, but to further investigate this I control for the four most dominating crops grown in 2009. As these crops are annual crops they would be planted in the rainy season in the second quarter of 2009 and hence, I restrict my sample of adopting farmers to those who adopted later than June 2009. If all farmers who grew traditional maize were connected and also converted to banana cultivation because bananas were more profitable than traditional maize, then the dummy for traditional maize cultivation would take the explanatory power away from the number of banana growers in the network. Results presented in column (3) of Table 6.4 do not support this story. The parameter for the RIPAT network is still significant at the one percent level and not more than one standard error lower than the result from the baseline specification. The dummy variables for cultivation of improved and traditional maize, beans and vegetables are jointly insignificant with a p-value of 0.697.[43] Hence, the results suggest that the network effects found are not driven by network formation based on previously grown crops.

In column (4) I study the puzzling result of a negative impact of the size of the information network conditional on the number of banana growers in the network. An explanation could be that farmers with large networks are mainly preoccupied with other activities than farming, such as transport or petty trade, which implies many contacts in the local community and less time for laborious farming activities. I hence include a binary variable taking the value one if the most important income source in the household is *not* agriculture or livestock keeping (i.e. the most important income source is either petty trade, transport, casual or wage labour, transfers or other income generating activities). As expected, this variable is negatively associated with adoption but the estimates of the network variables are unchanged.

Finally, I investigate whether the self-selection mechanism into RIPAT could be driving the network effects. A farmer who is initially interested in the new banana cultivation could choose not to participate in RIPAT if she knew several people who signed up for the project and who could give her seedlings and teach her the planting techniques. As participation in a RIPAT group requires time consuming tasks such as cultivation of the

[43]Wald test of joint significance: $\chi^2_{(4)} = 2.21$, $p = 0.697$.

common plot and participation in weekly meetings, this strategy could be optimal for farmers who face constraints on time or household labour. It has the flavour of the strategic delay found in the 'target input' model where farmers would delay adoption if they know enough adopters in order to avoid costly experimentation. Obviously, self-selection *out* of RIPAT and subsequent adoption through the information network would also constitute a social learning effect, but from a policy perspective this effect is of a smaller interest than the case where the perceived profitability of new technologies disseminates through information networks. This is the case because if the network effects found are purely driven by self-selection out of RIPAT then the crux of the matter would be to convince the farmers of the profitability of a new technology at the initial village meeting. The interesting case is when social learning effects work through affecting the perceived profitability of the technology in networks. A smaller group of farmers experiment with a new technology and their experiences will subsequently affect the perception in the local community of the relative profitability of the applied technology through their information networks. In this case, projects could simply target well-connected individuals in local communities. Through experimentation in the local circumstances, they would form a perception of the profitability of the new technology. This perception would then disseminate in their large networks increasing the reach of the project.

Thus, I find it important to distinguish the two different channels. I do this by splitting the sample of adopting farmers on early and late adopters, where early adopters planted their first improved bananas in 2006 or 2007. Within these two years RIPAT farmers would have time to plant bananas on their own farm and the plants would grow sufficiently to produce seedlings that can be shared in the network. I assume that the type of farmers who self-select out of RIPAT because they have network members in a RIPAT group would adopt as soon as possible and hence they would fall in the category of early adopters. If they do believe initially that banana cultivation is more profitably than other crops they grow, the optimal strategy would be to adopt as soon as possible. Column (5)-(6) in Table 6.4 presents the results for early and late adopters, respectively. There is indeed a larger parameter estimate for the RIPAT network effect among the early adopters providing support of this special type of strategic delay. Nevertheless, the estimate for late adopters is less than a standard error smaller than the baseline result and still significant at the one percent level. So even though some of the learning effects found in this paper may be explained by self-selection out of RIPAT groups, it appears that social learning also takes place through the perception of the relative profitability of banana cultivation.

7 Placebo study results

Is the strong correlation between RIPAT farmers in network and adoption behaviour simply driven by correlation in farmer specific characteristics within networks? If two entrepreneurial farmers discuss farming issues with each other, a correlation in their adoption behaviour may be due to the fact that they are both willing to experiment rather than social learning taking place. I address this question by considering the adoption of other cash crops and how that depends on the number of RIPAT farmers in the network. Table 7.1 presents the estimates from logistic regressions of adoption of vegetables, sunflowers and sugarcanes, respectively.

A positive correlation between the number of RIPAT banana growers in network and adoption of vegetables could indicate that unobserved farmer characteristics were correlated, but if the RIPAT banana growers in the network also grew vegetables a positive correlation would occur in the case of social learning within vegetables production. Hence, I control for the number of RIPAT banana growers in network who *also* grow the placebo crop. Since only 13 percent of the farmers grow sunflowers and eight percent grow sugarcane there are several subvillages with no variation in the adoption of the placebo crop which results in a fewer number of observations.

The results in Table 7.1 show that the number of banana growers in network cannot explain adoption of any of the three placebo crops. The lower number of observations in particular in column (2) and (3) raises the question of whether the models have any power to explain the variation in adoption of sunflower and sugarcane. If the effects of network

Table 7.1: Placebo study results, adoption of vegetables, sunflower and sugarcane

	(1) Vegetables	(2) Sunflower	(3) Sugarcane
RIPAT banana growers in network	0.313	0.213	-0.0552
	(0.256)	(0.203)	(0.248)
Non-RIPAT banana growers in network	0.114	-0.00148	0.0317
	(0.185)	(0.157)	(0.185)
Network size	0.0494	0.0295	-0.0621
	(0.0342)	(0.0741)	(0.0994)
RIPAT growing vegetables	-0.276		
	(0.343)		
RIPAT growing sunflowers		0.241	
		(0.578)	
RIPAT growing sugarcanes			0.747
			(0.914)
Observations	498	340	310
Adjusted pseudo R^2 [a]	0.116	0.102	0.081

Notes: Conditional logit coefficient estimates accounting for subvillage fixed effects. Standard errors in parentheses. Farmer and household characteristics are included in all specifications, but in column (2) religion dummies are excluded as Catholic dummy predicts non-adoption perfectly leading to drop of 43 observations. Results are robust to inclusion of religion dummies. The number of crops grown in 2010 is subtracted traditional and improved bananas and the placebo crop. Due to lack of variation in adoption within some subvillages, the number of observations is lower than 520.
[a] Adjusted pseudo R^2 is calculated as $1 - {(logL_m - K)}/{logL_0}$ where $logL_m$ refers to the log likelihood of the specified model with K variables, and $logL_0$ is the log likelihood of a model with only a constant.

variables are not significantly different from zero simply due to large standard errors caused by smaller sample sizes then the placebo test has no bite. Exploiting variation across subvillages in order to keep more observations in the sample would not resolve the issue: if bananas and vegetables require the same growing conditions adoption will be correlated within subvillages and if farmers discuss farming issues with their neighbours, then the correlation between RIPAT banana growers in network and adoption of vegetables would not give evidence of correlated farmer specific characteristics. Instead, I start out by regarding the estimated logit coefficients. To make an exact comparison with the baseline results would require calculation of marginal effects which is not feasible. Nevertheless, I note that the estimated parameter for RIPAT banana growers in network is negative with respect to adoption of sugarcanes and that it is three to five times smaller than the baseline estimate for vegetable and sunflower adoption, respectively. Standard errors are less than 1.5 times larger in the baseline regression. This suggests that the insignificance of the parameters is not purely driven by an increase in the standard errors. Accordingly, the parameters for non-RIPAT banana growers in network are also less than half of the baseline estimate while standard errors are 1.8 times larger.

To give an imprecision of the explanatory power of the models, I have included the adjusted pseudo R-squared in Table 7.1.[44] For comparison, the adjusted pseudo R-squared of the baseline model is 0.156. As the placebo regressions have R-squares of 0.081-0.116, they do not appear to explain much less than the baseline model.[45]

Even though none of the network variables have any explanatory power, some of the other covariates are significantly correlated with adoption of the placebo crops (results not reported). In particular the number of crops grown in 2010, net of banana cultivation and the given placebo crop, is positively and significantly associated with the adoption of all three placebo crops. The same was the case for adoption of banana cultivation suggesting that this variable captures the entrepreneurial spirit of the farmer or the "willingness to experiment". Indeed, if I exclude the number of crops grown in 2010 from the placebo regressions, the parameter for banana growers in network is significant at a ten percent level for the adoption of vegetables. Hence, there may be some correlation in farmer characteristics such as the "willingness to experiment", but this is to a large degree captured in the number of crops grown in 2010.

This placebo study further supports the existence of social learning since it does not provide support for a potential correlation in unobserved farmer characteristics within networks as the main driver of the network effects found.

[44]It should be noted that the pseudo R-squared can *not* be interpreted as the share of the variation of the outcome variable explained by the model. However, it can give an indication of the explanatory power of the model.

[45]As I do not include the number of adopters of the placebo crops in the placebo regressions, it could be interesting to compare the adjusted pseudo R-squared to that of the baseline model excluding the two variables capturing banana growers in network. That yields an adjusted pseudo R-squared of 0.069 which is even lower than the R-squared of the placebo regressions.

8 Conclusion and discussion

This empirical paper studies the role of information networks in the diffusion of improved banana cultivation in northern rural Tanzania. The new cultivation technique was introduced by an agricultural project, and the results show that discussing farming issues with banana growing farmers who participate in the project increases the likelihood of adopting improved banana cultivation. Furthermore, discussing farming issues with other banana growers also seems to have an impact on adoption, though this effect is weaker and less robust.

It is difficult to identify social interaction effects in the absence of experimental variation in the data. The issue is treated rigorously in section 5 of the paper. A test is suggested for the potential reverse causality from the ex ante perceived profitability of banana cultivation to network formation. Results indicate no presence of reverse causality in the data. Whether the network effects found are driven by contextual effects (such as the wealth of network members) is investigated by splitting the project participants in networks by different socioeconomic characteristics. The network effects do not appear to be heterogeneous across the characteristics of the project farmers in the network suggesting that the results are not driven by contextual effects. Confounding factors such as growing conditions and institutions are accounted for by including the number of banana growers within half a kilometre and by only considering variation within subvillages. Finally, the potential role of unobserved farmer-specific characteristics is examined through a placebo study. The network effects found may be upward biased if entrepreneurial farmers know more banana growers but also are more likely to adopt banana cultivation regardless of their network. If the number of banana growers in network proxies for the entrepreneurial skills of the farmer, it should also be positively correlated with the adoption of other lucrative cash crops. I examine the adoption of vegetable, sunflower and sugarcane production, but do not find the network variables to have any explanatory power in these regressions. Hence, I conclude that unobserved characteristics do not seem to drive the network effects found within banana cultivation.

Generally, I find that the impact of discussing farming issues with project participants on adoption of banana cultivation is large and the network effect is very robust across specifications. Furthermore, none of the identification tests are in disfavour of the presence of endogenous social interaction effects. Though I cannot perfectly rule out all issues of identification, the analysis clearly points to the existence of social learning in the context under study.

This finding adds to the thin empirical literature on social learning within adoption of agricultural technologies in developing countries. It confirms the importance of social networks as concluded by Conley and Udry (2010) and Bandiera and Rasul (2006). Furthermore, I contribute to the literature on adoption of technology by studying the diffusion of technology beyond project participants of an agricultural project. To my knowledge, this has not been done before. It is important for policy implications to study this kind of second generation take-up because it can provide advice on whom to target for agricultural projects in order to obtain the largest possible impact. The results presented in this paper suggest that targeting well-connected farmers will increase the impact of a similar future project through diffusion of technology in the local community.

Bandiera and Rasul (2006) find that the adoption of sunflower cultivation has an inverse U-shaped relationship to the number of adopters in the network. They argue that the negative marginal effect of knowing an extra adopter when the farmer knows ten or more adopters can be interpreted as evidence of strategic delay. If information is shared in the network then knowing many adopters allows the farmer to free-ride on their experimentation and hence induces her to delay her decision to adopt and avoid the costly experimentation. I also find an indication of an inverse U-shaped relationship between adoption of banana cultivation and project banana growers in network with a peak of three adopters in network. However, 99 percent of my sample knows three or fewer project participants who grow improved bananas. Hence, this result does not allow me to conclude that farmers are strategically delaying their adoption decision. But since participation in the project under study was voluntary that allows me to revise the strategic delay found in Bandiera and Rasul (2006), though in another dimension. A farmer knowing several other farmers who signed up for the project may deliberately have chosen not to sign up in expectation of future social learning effects. Such a farmer would then receive improved banana seedlings and advice from participating farmers as soon as they had learned the new technique. This strategy could be optimal for a farmer who has a high opportunity cost of labour because project participation required presence at weekly meetings and cultivation of a common plot. Since this type of strategic delay would imply adoption of improved banana cultivation within two years after project start, I study if there are differential network effects for early and late adopters. I do indeed find that network effects are stronger for early adopters suggesting the presence of this special type of strategic delay, but the strong average network effect found among all adopters persist among the late adopters. The latter indicates that the network effects cannot be fully explained by self-selection out of the project for well-connected farmers, which is important for policy implications. Though farmers exerting strategic delay do indeed rely on social learning, this type of social learning does not indicate that a newly introduced technology would diffuse in the surrounding community. On the other hand, an interpretation of the network effect for the late adopters is that experimentation among their network members have affected their perception of the relative profitability of the new technology and thereby induced them to adopt it. This kind of social learning results in a multiplier effect of an agricultural project that succeeds in introducing a new technology which is more profitable than existing technologies and at the same time compatible with local norms and circumstances.

The results in the paper reveal a larger and more robust effect of knowing banana growing project participants than of knowing other banana growers. This can be interpreted in different ways. It could be an indication that the learning effects subside as we move further away from project participants. This implies that the multiplier effect of the project is limited to the network of project participants and we should only expect a small degree of diffusion from adopting non-participants and further on in their networks. On the other hand, the smaller effect of banana growing non-participants in the network may be a pure product of timing: Everything else equal, I expect banana growers who are not part of the project to have planted later than project participants. Hence, they may not yet have harvested their first banana bunch and hence, their network members may not be able to deduce the profitability of banana cultivation yet. In this case, the differential network effects do not contradict that the diffusion of knowledge continues beyond the second generation take-up. Finally, the specific design of the project pertaining the diffusion of technology may play a role. Participating farmers were obliged to pass on banana seedlings thrice which may explain why knowing a project participant has a stronger impact on adoption than knowing another banana grower. Nevertheless, I find that network effects persist if I limit my sample of adopting farmers to those who actively asked for banana seedlings. This suggests that the network effects are not purely driven by project design.

Whether the social learning effects found in this paper will persist and cause improved banana cultivation to further diffuse in the community remains an open question.

9 Bibliography

Bandiera, O., and I. Rasul (2006): “Social Networks and Technology Adoption in Northern Mozambique,” *The Economic Journal*, 116, 869–902.

Banerjee, A. V. (1992): “A Simple Model of Herd Behavior,” *The Quarterly Journal of Economics*, 107(3), 797–817.

Bardhan, P., and C. Udry (1999): *Development Microeconomics*. Oxford University Press.

Berry, D., and B. Lindgren (1996): *Statistics: Theory and methods*. Duxbury Press.

Besley, T., and A. Case (1993): “Modeling Technology Adoption in Developing Countries,” *The American Economic Review*, 83(2), pp. 396–402.

——— (1994): “Diffusion as a Learning Process: Evidence from HYV Cotton,” Working Papers 228, Princeton University, Woodrow Wilson School of Public and International Affairs, Research Program in Development Studies.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): “Identification of peer effects through social networks,” *Journal of Econometrics*, 150(1), 41 – 55.

Brock, W. A., and S. N. Durlauf (2001): “Chapter 54: Interactions-based models,” in *Handbook of econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5, pp. 3297 – 3380. Elsevier.

——— (2007): “Identification of binary choice models with social interactions,” *Journal of Econometrics*, 140(1), 52 – 75.

Chamberlain, G. (1982): “Analysis of Covariance With Qualitative Data,” Working Paper 325, National Bureau of Economic Research.

Conley, T. G., and C. R. Udry (2010): “Learning about a New Technology: Pineapple in Ghana,” *The American Economic Review*, 100(1), 35–69.

Deitchler, M., T. Ballard, A. Swindale, and J. Coates (2011): "Introducing a Simple Measure of Household Hunger for Cross-Cultural Use," Discussion Paper Technical Note No. 12, Food And Nutrition Technical Assistance, US Aid, available online at http://www.fantaproject.org/downloads/pdfs/TN12_HHS.pdf, accessed 2 March, 2012.

Duflo, E., and E. Saez (2002): "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," Working Paper 8885, National Bureau of Economic Research.

Durlauf, S. N., and Y. M. Ioannides (2010): "Social Interactions," *Annual Review of Economics*, 2(1), 451–478.

Ellison, G., and D. Fudenberg (1993): "Rules of Thumb for Social Learning," *Journal of Political Economy*, 101(4), pp. 612–643.

Evenson, R. E., and L. E. Westphal (1995): "Chapter 37 Technological change and technology strategy," in *Handbook of Development Economics*, ed. by J. Behrman, and T. Srinivasan, vol. 3, Part A, pp. 2209 – 2299. Elsevier.

Feder, G., R. E. Just, and D. Zilberman (1985): "Adoption of Agricultural Innovations in Developing Countries: A Survey," *Economic Development and Cultural Change*, 33(2), pp. 255–298.

Foster, A. D., and M. R. Rosenzweig (1995): "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103(6), pp. 1176–1209.

——— (2010): "Microeconomics of Technology Adoption," *Annual Review of Economics*, 2(1), 395–424.

Gausset, Q. (forthcoming): "Local adoption of social and agricultural technologies," in *Farmers' Choice: Evaluating an approach to agricultural technology adoption*, ed. by H. B. Lilleør, and U. Lund-Sørensen. Practical Action Publishing.

GLEWWE, P., AND M. E. GROSH (2000): *Designing household survey questionnaires for developing countries, lessons from 15 years of the living standards measurement study*, vol. 1-3. Oxford University Press: World Bank, Washington D.C.

GRANOVETTER, M. S. (1973): "The Strength of Weak Ties," *American Journal of Sociology*, 78(6), pp. 1360–1380.

GRILICHES, Z. (1957): "Hybrid corn: An exploration in the economics of technological change," *Econometrica*, 25(4), 501–522.

HIJMANS, R., S. CAMERON, J. PARRA, P. JONES, AND A. JARVIS (2005): "Very high resolution interpolated climate surfaces for global land areas," *International journal of climatology*, 25(15), 1965–1978.

JOVANOVIC, B., AND Y. NYARKO (1994): "Learning by Doing and the Choice of Technology," Working Paper 4739, National Bureau of Economic Research.

KREMER, M., AND E. MIGUEL (2007): "The Illusion of Sustainability," *The Quarterly Journal of Economics*, 122(3), 1007–1065.

MADDALA, G. S. (1991): "A Perspective on the Use of Limited-Dependent and Qualitative Variables Models in Accounting Research," *The Accounting Review*, 66(4), pp. 788–807.

MAGUZU, C., D. RINGO, AND J. VESTERAGER (forthcoming): "Presentation of RIPAT - Core Components and Project Implementation," in *Farmers' Choice: Evaluating an approach to agricultural technology adoption*, ed. by H. B. Lilleør, and U. Lund-Sørensen. Practical Action Publishing.

MANSKI, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60(3), 531–542.

——— (2000): "Economic Analysis of Social Interactions," *The Journal of Economics Perspectives*, 14(3), 115–136.

——— (forthcoming): “Identification of treatment response with social interactions,” *The Econometrics Journal*.

Manski, C. F., and S. R. Lerman (1977): “The Estimation of Choice Probabilities from Choice Based Samples,” *Econometrica*, 45(8), pp. 1977–1988.

McFadden, D. (1973): “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Frontiers of Econometrics*, ed. by P. Zarembka, pp. 105–142. Academic Press.

Mogensen, H. O., and E. K. Pedersen (forthcoming): “Household Dynamics and Gender Politics: Female farmers in RIPAT 1,” in *Farmers' Choice: Evaluating an approach to agricultural technology adoption*, ed. by H. B. Lilleør, and U. Lund-Sørensen. Practical Action Publishing.

Munshi, K. (2004): “Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution,” *Journal of Development Economics*, 73(1), 185 – 213.

——— (2007): “Chapter 48 Information Networks in Dynamic Agrarian Economies,” in *Handbook of Development Economics*, ed. by T. P. Schultz, and J. A. Strauss, vol. 4, pp. 3085 – 3113. Elsevier.

O'Mara, G. (1971): “A decision-theoretic view of the microeconomics of technique diffusion in a developing country,” Ph.D. thesis, Department of Economics, Stanford University.

Prentice, R. L., and R. Pyke (1979): “Logistic Disease Incidence Models and Case-Control Studies,” *Biometrika*, 66(3), pp. 403–411.

Rogers, E. (1995): *Diffusion of innovations*. Free Press, New York, 4 edn.

Schreiner, M. (2011): “A Simple Poverty Scorecard for Tanzania,” Discussion paper, Progress out of Poverty, Grameen Foundation, available online at http://progressoutofpoverty.org/tanzania, accessed 2 March, 2012.

Scott, J., and P. Carrington (2011): *The SAGE Handbook of social network analysis*. Sage Publications Ltd.

Sunding, D., and D. Zilberman (2001): "Chapter 4 The agricultural innovation process: Research and technology adoption in a changing agricultural sector," in *Agricultural Production*, ed. by B. L. Gardner, and G. C. Rausser, vol. 1, Part A of *Handbook of Agricultural Economics*, pp. 207 – 261. Elsevier.

Van den Broeck, K., and S. Dercon (2007): "Social Interactions in Growing Bananas: Evidence from a Tanzanian Village," The Centre for the Study of African Economies Working Paper Series, Working Paper 264, Oxford University.

A Appendix - Data collection

The data collection took place in January 2011, were 2,639 households were interviewed in total. The survey was funded by the Rockwool Foundation, designed by a Rockwool Foundation Research Team comprising Helene Bie Lilleør (HBL), Head of Evaluation, and I, administered by RECODA (the NGO that also implemented RIPAT) and supervised by Economic Development Initiative (EDI), which is a Tanzanian survey and consultancy house. The main purpose of the data collection was to perform an impact evaluation of RIPAT, and for this purpose all RIPAT households were interviewed together with a corresponding number of control households from other villages. In addition, non-participating households in RIPAT villages were interviewed for the purpose of studying the dissemination of banana cultivation. First, I will describe the composition of the sample comprising randomization and stratification issues and then follows a description of the development of survey instruments and the implementation of the data collection.

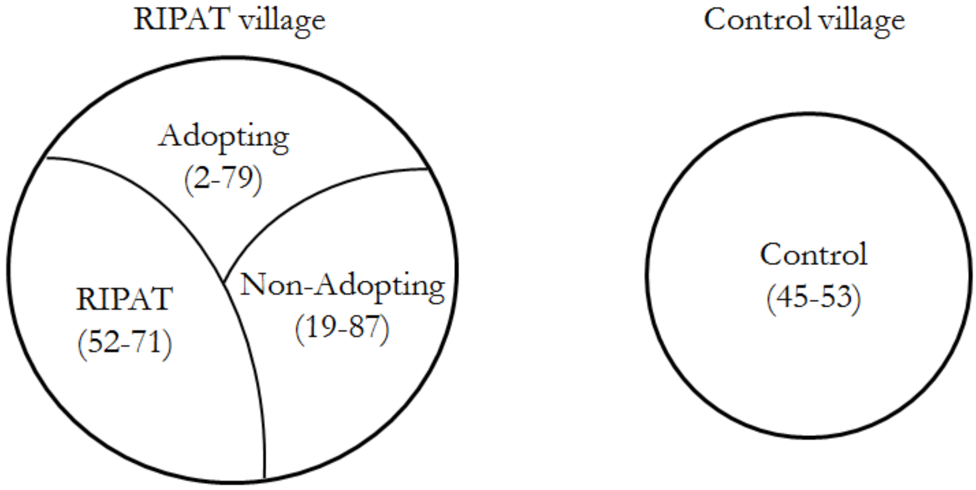
Table A.1: Sample of households

	RIPAT 1	RIPAT 3
RIPAT	506	564
Control	490	482
Adopting	224	-
Non-adopting	373	-
Total	1,593	1,046

Sample composition

The survey covered four different types of households: *RIPAT* households, *control* households in control villages, and two types of non-RIPAT households in RIPAT villages, namely *adopting* and *non-adopting* households that were categorized depending on whether or not they grew bananas. The number of households interviewed is presented in Table A.1. The first two groups of households where interviewed in order to perform the impact evaluation. RIPAT and control households were interviewed both from the Arumeru District where RIPAT1 took place but also in Karatu District where RIPAT3 took place.

Figure A.1: Sample composition



Note: Number of observations per village in parentheses

Data collection in RIPAT2 was abolished because of cultural difficulties in collecting quantitative data among the Masai people. Adopting and non-adopting households were only interviewed in RIPAT1 villages as the dissemination of banana cultivation was still very sparse in the RIPAT3 villages at the time of the data collection preparations. An overview of the sample composition is depicted in Figure A.1.

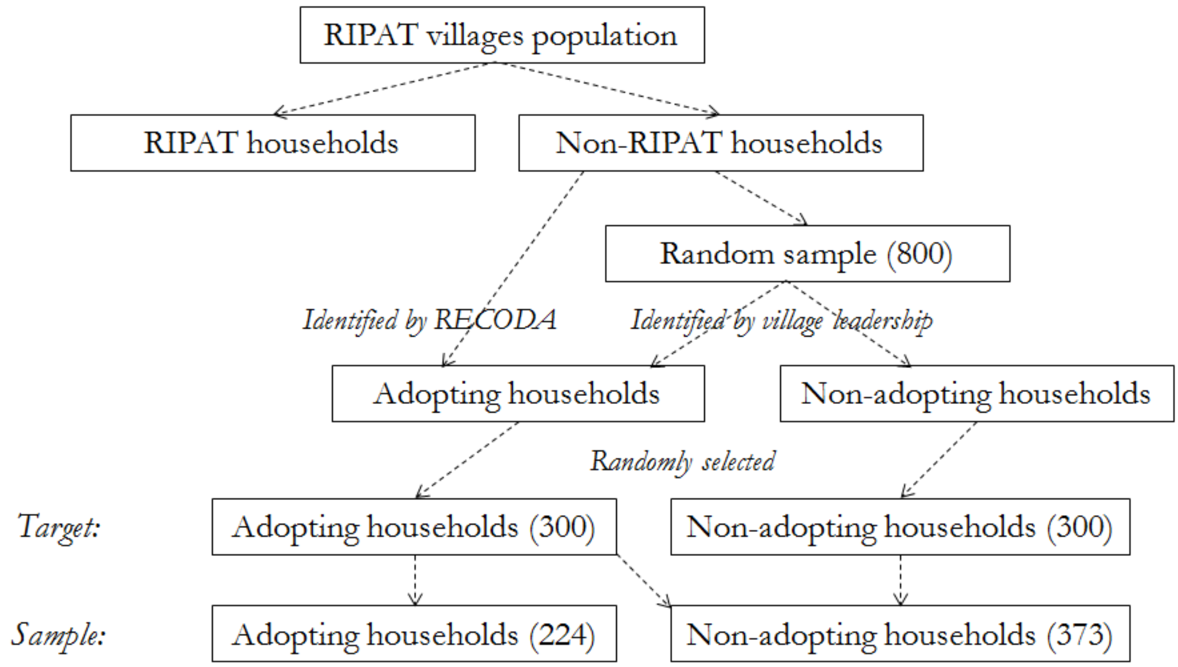
RIPAT households

Based on initial enrollment list and visits to RIPAT groups, every farmer who had ever been a member of a RIPAT group was identified and included in the survey sample. Almost all (95.1 percent) of the RIPAT households completed an interview.

Adopting and non-adopting households

In RIPAT villages, non-RIPAT households were sampled mainly for the purpose of this study. RECODA staff collected village lists from all eight RIPAT villages consisting of a complete list of all households in the village. RIPAT households were identified from the list, and 100 non-RIPAT households were selected at random from each village. Randomization was implemented by numbering the households on the village lists and let Excel draw random numbers among the number of households in the village. RECODA

Figure A.2: Sample selection



had a list of households who had received banana seedlings from RIPAT farmers and these households were identified from the village lists and added to the random sample if they were not already listed. For this random sample and additional adopting households, RECODA staff collected information from the local leadership on whether the household was farming and whether the household grew improved bananas. This was done to avoid non-farming households in the sample and to confirm whether a household was adopting or not. On the basis of this information, I stratified the sample to obtain an even number of adopting and non-adopting households. This was done in order to ensure that the final sample would include enough adopting households which are the households of interest for this study. Had I simply drawn a random sample, some of the villages would only have one or two adopting households in the sample which would impede the analysis of the adoption decision. In that case, single observations would be pivotal for the outcome of the analysis which is undiserable. Figure A.2 shows a chart of the sample selection procedure for the adopting and non-adopting households.

The degree of adoption varied considerably across villages, so I decided not to have the same sample size of non-adopting households in each village. The budget covered in-

Table A.2: Sample of non-RIPAT households

	Adopting households			Non-adopting households
Village	Identified	Target	Sample	Sample
Kwaugoro	94	79	80	89
Karangai	33	27	20	37
Kikwe	15	15	7	19
Maweni	27	23	22	24
Majimoto	63	53	28	70
Mungushi	15	15	3	32
Manyata	41	34	14	50
Marurani	65	54	50	52
Total	353	300	224	373

terviews of 600 non-RIPAT households implying a *target* of 300 interviews with adopting households. From the information of the local leaders I listed the number of adopting households identified in a given village as can be seen in the first column of Table A.2 and calculated their share out of the total number of identified adopting households. I then let the sample of households in this village constitute the same share out of the sample of non-RIPAT households. E.g. 94 adopting households were identified in Kwaugoro out of 353 in total, that is 26.6 pct. So 26.6 pct of the target sample of 300 adopting households should come from Kwaugoro, corresponding to a target of 80 adopting households. However, HBL and I decided on a minimum of 30 households in each village (15 adopting households) as we also wanted to exploit the information on non-RIPAT farmers for the impact evaluation, and for that purpose the sample of non-RIPAT farmers in each village should not be too small. This implied that the target samples in two villages (Kikwe and Mungushi) were adjusted upwards and the samples in the remaining villages were adjusted downwards accordingly. The target samples are presented in column two of Table A.2.

During the data collection, several issues made the final sample deviate from the target. One important explanation is that 73 households showed up to be non-adopting although the village leadership and RECODA records had stated otherwise and only 48 expected non-adopting households turned out to be adopting. This explains a difference of 50 households between the samples of adopting and non-adopting households. In addition, unavailable adopting households were supposed to be replaced by another adopting household to the extent possible. But in two villages (Karangai and Mungushi) no more adopting farmers could be identified, so unavailable adopting households were instead replaced by non-adopting households. This further explains a difference of 38 households between the two categories. Also, village leaders were not able to categorise 45 households a priori (21 of these were in Manyata) which hampered targeting of households. Furthermore, the replacement rules were not abided in the case of Majimoto and several non-adopting households were interviewed instead of targeted adopting households.

Control households

We chose control villages with the help of the District Offices in Arumeru and Karatu District. They appointed 10 and 12 villages, respectively, that were similar to RIPAT villages with respect to a range of characteristics, in particular size, agroclimatic zone, water supply, main religion and ethnicity, schools and health facilities, and distance to nearest road and market. From the chosen control villages RECODA staff collected village lists which consisted of a complete list of all households in the village. Based on power calculations and budget restrictions, HBL and I decided to have 480 control households in each district implying 48 control households in each control village in Arumeru district and 40 control households in each control village in Karatu District.

I stratified the sample of control households by subvillage since RIPAT households were purposely chosen from all subvillages in the RIPAT villages. To ensure that all subvillages where represented according to size, I calculated the share of households in a subvillage out of the total number of households in the village and let the subvillage have the same share in the random sample.

Female headed households were overrepresented in the RIPAT groups as a result of the policy to have at least 50 percent female farmers in the groups. It was noted on the initial village lists whether a household was female headed so that information was used to stratify the final random sample of control households such that they had the same share of female headed households as the RIPAT groups.

RIPAT households where restricted to be farming households, to have at least one acre and not more than five acres of land. Since the upper limit was not followed very strictly, we chose to expand the upper limit to eight acres of land. To ensure that the control farmers would be comparable to RIPAT farmers, an initial random stratified sample of 100 households was drawn for which RECODA staff gathered information from the local leadership on whether the listed households were farming and had acres within the range. Based on this information some households were dropped from the sample. Furthermore, the first questions in the questionnaire referred to whether the household is engaged in farming activites and the number of acres they have, and if a control household did not fulfil the criteria it was replaced. In addition, the questionnaire included a brief presentation of an NGO project similar to RIPAT asking if the farmer would want to participate if possible. If the farmer answered no, the household was also replaced.

Development of survey instruments

The survey consisted of three questionnaires: a household questionnaires, a village questionnaire and RIPAT group questionnaire.

I designed the household questionnaire such that it could be administered to all four types of households with an elaborate skip pattern allowing some sections to be administered only to one type of households. The questionnaire included the sections listed in Table A.3.

Sections C, F, H, K and O are inspired by questionnaires from the Living Standard Measurement Surveys (Glewwe and Grosh, 2000). Section A and L are adapted from the household questionnaire of the Karonga Assessment of Vulnerability in Malawi, conducted by the RFF in 2010. Numeracy questions in section F and the risk game in section G are taken from a survey on sexual and reproductive health in Uganda completed by Innovations for Poverty Action, kindly made available by Julian Jamison, Federal Reserve, Bank of Boston. Section I and M are copied from the Household Budget Survey 2006/07 conducted by the National Bureau of Statistics, Tanzania, in order to enable the construction of a poverty score card developed from the 2006/07 survey. The last part of section N is developed by US Aid to measure household hunger (Deitchler, Ballard, Swindale, and

Table A.3: Sections of household questionnaire

Section	**Content**
Section A	Administrative data and GPS measures of the household dwelling
Section B	Identifying type of household and main respondent
Section C	Crop cultivation and land holdings
Section D	Banana cultivation (RIPAT and adopting farmers only)
Section E	Information network (adopting and non-adopting farmers only)
Section F	Farmer characteristics: religion, ethnicity and numeracy
Section G	Risk game
Section H	Shocks and participation in other projects
Section I	Household assets
Section J	Livestock
Section K	Household roster: age, education, marital status and health
Section L	Income sources
Section M	Household facilities: quality and remoteness of housing
Section N	Food security
Section O	Measurement of children below six years

Coates, 2011). Finally, I have developed section B, D, E, J and the first part of section N for the purpose of this survey, where information network questions in section E are inspired by network measures used in Conley and Udry (2010) and Bandiera and Rasul (2006).

The household questionnaire can be divided into three parts: section C-G where the preferred respondent is the *farmer*, section H-L where the preferred respondent is either the *farmer* or the *adult female* in the household, and section M-O where the preferred respondent is the *adult female*. The preferred respondents are chosen this way to get the most reliable information both on farming activities and household facilities and food security. The *farmer* is identified in section B as either the member of the RIPAT group in RIPAT households; the farmer who "decided to grow improved bananas" in adopting households; or the person "who has taken most decisions for farming activities" in non-adopting households in RIPAT villages and control households in control villages. This manner of identifying the main respondent was chosen because this allowed us to collect personal information on the decision maker of interest.

The village questionnaire is highly inspired by the village questionnaire from the Karonga Assessment of Vulnerability in Malawi, adapted to local circumstances and the purposes of the data collection. It was administered as a group interview with the Village

Executive Officer, the Village Chairperson and the Village Agricultural Extension Officer or a knowledgeable farmer from the village if the extension officer was not available. The virtue of a group interview with several key informants is that they can discuss and confirm the answers and hence, we elicit the most precise and trustworthy information.

The RIPAT group questionnaire was developed for the purpose of this survey and administered to two persons out of the group triumvirate: the leader, secretary and accountant, immediately after a household interview with one of them.

Implementation of the data collection

The survey team

The data collection was funded by the Rockwool Foundation and led by HBL and I, with assistance from Maria Fibæk, an economics master student at the University of Copenhagen. It was implemented in collaboration with RECODA, the implementing organization of RIPAT, who identified six supervisors out of which four were permanent RECODA staff. One of the permanent RECODA staffers was chosen to supervise the data entry process, while two others were in an advanced state of pregnancy and hence unable to do extensive field work. As five field supervisors were needed, an external supervisor was hired together with an additional supervisor among the applicants for the enumerator jobs. This implies that only one in five field supervisors was a RECODA staffer. We hired 25 enumerators to enter in teams of five with one supervisor per team. In the hiring process we emphasised that enumerators should not be associated with RECODA in order to avoid that a RIPAT associated enumerator would affect RIPAT farmers to respond more positively to subjective questions (also known as the Hawthorne effect). Initially, we wanted to hire a local survey manager to administer the survey, but the person we had found for the job bailed out in the last minute, and instead we engaged a survey expert and a field manager from EDI. In addition, the vice president of RECODA was responsible for practical matters with respect to the data collection such as transport, lunch and printing of questionnaires among many other things.

Preparations for the data collection

Prior to data collection, the supervisors collected village lists, then stratification and randomisation took place, and the supervisors gathered additional information from the local village leaderships on the sample of households as described under *Sample composition*. All three questionnaires were pilot-tested by the supervisors guided by Maria Fibæk and I. In particular the household questionnaire was subject to extensive pilot-testing of all parts of the questionnaire to ensure that response categories were adequate, that questions were easy to understand and induced consise responses. During pilot-testing, Maria and I trained supervisors in interview techniques, checking questionnaires and organising field work. All questionnaires were translated into Swahili by the group of supervisors and translation was discussed among them and with us to ensure the correct intention of the question. Even during training of enumerators and the first week of the actual data collection minor adjustments of the questionnaires were made based on experiences from enumerators and supervisors.

The survey expert and the field manager from EDI, together with Maria Fibæk, were responsible for the training of the enumerators comprising class room presentation of the questionnaires and training in interview techniques, combined with pretesting of questionnaires in villages that were not part of the sample.

The data collection

The actual data collection took place from January 3rd to February 4th, 2011. It was implemented by the team of supervisors and enumerators, organised by the field manager and supervised by the survey expert who was in the field half of the time. Upon completion of a questionnaire, the enumerator would first check the questionnaire for mistakes, then the supervisor would go through the questionnaire to identify inconsistencies, wrong skip patterns or the like. In addition, five percent of the questionnaires were also checked by the field manager or an assistant supervisor at the RECODA office. Furthermore, two percent of the questionnaires were subject to back-checks, where the field manager or supervisor returned to an interviewed household and repeated parts of the interview to ensure that the enumerator had not invented the answers in the questionnaire.

Data entry

Once the final questionnaires were ready, they were coded into CSPro (Census and Survey Processing System) which is a software package for entering survey data developed and designed by the United States Census Bureau. Cathrine Søgaard Hansen, a research assistant at the Rockwool Foundation Research Unit, programmed this CSPro template under my supervision. She also led the data entry process in Tanzania and assisted the RECODA data entry supervisor. Ten local data entry clerks were hired to enter the information from the questionnaires into the CSPro template. All questionnaires were entered twice by two different data entry clerks and in case of disagreement between the two entries both clerks would make corrections to their entry and the entries would be compared again. An entry was accepted when there was no longer a disagreement between the two entries. For practical purposes the data entry clerks were teamed up in pairs with that unfortunate implication that one pair colluded and did not enter all the data from the questionnaires. However, this was found out and corrected.

B Appendix - Choice-based sampling in a logit model

This appendix shows that the logit model provides consistent estimates of the parameters in the case of choice-based sampling.[46]

Assume that the probability of adoption in the population, $\tilde{P}(a_i = 1)$,[47] is logistically distributed and depends on a range of farmer and household characteristics, Z_i and subvillage fixed effects, α_s:

$$\tilde{P}(a_i = 1|Z_i, \alpha_s) = \Lambda(\theta Z_i + \alpha_s) = \frac{exp(\theta Z_i + \alpha_s)}{1 + exp(\theta Z_i + \alpha_s)} \tag{B.1}$$

For simplicity assume that all the covariates are discrete, such that we can consider

[46] I would like to thank Professor Bo Honoré, Princeton University, for indispensable help with the following derivation.

[47] I disregard the subscript t on adoption in this exposition.

probabilities instead of distributions. The result generalizes to continuous covariates.

The probability of adoption in the sample, $P(a_i = 1)$, conditional on covariates and subvillage fixed effects can be rewritten using Bayes rule:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{P(a_i = 1, Z_i, \alpha_s)}{P(Z_i, \alpha_s)}$$

We now use that the sample of adopting farmers is a random sample such that the probability of the covariates given that the farmer is adopting is the same in the sample and in the population, $P(Z_i, \alpha_s|a_i = 1) = \tilde{P}(Z_i, \alpha_s|a_i = 1)$, and correspondingly for non-adopting farmers. In addition, we use the law of iterated expectations in the denominator:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{\tilde{P}(Z_i, \alpha_s|a_i = 1) \cdot P(a_i = 1)}{\sum \tilde{P}(Z_i, \alpha_s|a_i = 0) \cdot P(a_i = 0) + \sum \tilde{P}(Z_i, \alpha_s|a_i = 1) \cdot P(a_i = 1)} \tag{B.2}$$

Applying Bayes rule and using equation B.1, we can rewrite:

$$\tilde{P}(Z_i, \alpha_s|a_i = 1) = \frac{\tilde{P}(Z_i, \alpha_s, a_i = 1)}{\tilde{P}(a_i = 1)} = \frac{\tilde{P}(Z_i, \alpha_s) \cdot \Lambda(\theta Z_i + \alpha_s)}{\tilde{P}(a_i = 1)} \tag{B.3}$$

Correspondingly,

$$\tilde{P}(Z_i, \alpha_s|a_i = 0) = \frac{\tilde{P}(Z_i, \alpha_s) \cdot (1 - \Lambda(\theta Z_i + \alpha_s))}{\tilde{P}(a_i = 0)} \tag{B.4}$$

We now insert equation B.3 and B.4 in equation B.2:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{\frac{P(a_i=1)}{\tilde{P}(a_i=1)} \cdot \tilde{P}(Z_i, \alpha_s) \cdot \Lambda(\theta Z_i + \alpha_s)}{\frac{P(a_i=0)}{\tilde{P}(a_i=0)} \tilde{P}(Z_i, \alpha_s) \cdot (1 - \Lambda(\theta Z_i + \alpha_s)) + \frac{P(a_i=1)}{\tilde{P}(a_i=1)} \cdot \tilde{P}(Z_i, \alpha_s) \cdot \Lambda(\theta Z_i + \alpha_s)}$$

We divide numerator and denominator with $\tilde{P}(Z_i, \alpha_s)$ and insert the definition of the logistic distribution:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{\frac{P(a_i=1)}{\tilde{P}(a_i=1)} \cdot \frac{exp(\theta Z_i + \alpha_s)}{1+exp(\theta Z_i + \alpha_s)}}{\frac{P(a_i=0)}{\tilde{P}(a_i=0)} \cdot \frac{1}{1+exp(\theta Z_i + \alpha_s)} + \frac{P(a_i=1)}{\tilde{P}(a_i=1)} \cdot \frac{exp(\theta Z_i + \alpha_s)}{1+exp(\theta Z_i + \alpha_s)}}$$

Finally, we divide both numerator and denominator with the first term of the denominator and rearrange:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{exp(\theta Z_i + \alpha_s + ln(c))}{1 + exp(\theta Z_i + \alpha_s + ln(c))}, \qquad c \equiv \frac{P(a_i=1)/\tilde{P}(a_i=1)}{P(a_i=0)/\tilde{P}(a_i=0)} \tag{B.5}$$

Comparing the probability of adoption in the sample (equation B.5) with the probability of adoption in the population (equation B.1), it is evident that the choice-based sampling only affects the estimation of the subvillage fixed effects (or the constant in the case of no fixed effects) and hence, the estimated parameters of the covariates (θ) are unaffected by the sampling method.

C Appendix - Further robustness of results

Table C.1: Adoption of improved banana cultivation, robustness of results

	(1) Without network size	(2) Asked for seedling
RIPAT banana growers in network	0.857***	1.173***
	(0.158)	(0.190)
Non-RIPAT banana growers in network	0.0533	0.330***
	(0.0899)	(0.112)
Network size		-0.264***
		(0.0683)
Observations	514	472

Note: Conditional logit coefficient estimates accounting for subvillage fixed effects. Standard errors in parentheses. Farmer and household characteristics are included in all specifications. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.